

DEX-0259

BA

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
30 May 2002 (30.05.2002)

PCT

(10) International Publication Number
WO 02/42776 A2(51) International Patent Classification⁷: G01N 33/574(US). LIU, Chenghua [CN/US]; 1125 Ranchero Way #14,
San Jose, CA 95118 (US).

(21) International Application Number: PCT/US01/45654

(74) Agents: LICATA, Jane, Massey et al.; Licata & Tyrell
P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(22) International Filing Date:

1 November 2001 (01.11.2001)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/244,782 1 November 2000 (01.11.2000) US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA,
ZW.(63) Related by continuation (CON) or continuation-in-part
(CIP) to earlier application:US 60/244,782 (CON)
Filed on 1 November 2000 (01.11.2000)(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,
TG).(71) Applicant (*for all designated States except US*): DI-
ADEXUS, INC. [US/US]; 343 Oyster Point Boulevard,
South San Francisco, CA 94080 (US).

Published:

— without international search report and to be republished
upon receipt of that report

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): SUN, Yongming
[US/US]; 869 S. Winchester Boulevard, Apartment 260,
San Jose, CA 95128 (US). RECIPON, Hervé [FR/US]; 85
Fortuna Avenue, San Francisco, CA 94115 (US). CHEN,
Sei-Yu [CN/US]; 160 Mira Street, Foster City, CA 94404For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

WO 02/42776 A2

(54) Title: COMPOSITIONS AND METHODS RELATING TO PROSTATE SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic prostate cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating prostate cancer and non-cancerous disease states in prostate tissue, identifying prostate tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered prostate tissue for treatment and research.

44

DEX-0524

COMPOSITIONS AND METHODS RELATING TO PROSTATE SPECIFIC GENES AND PROTEINS

This application claims the benefit of priority from U.S. Provisional Application
5 Serial No. 60/244,782 filed November 1, 2000, which is herein incorporated by reference
in its entirety.

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and
10 polypeptides present in normal and neoplastic prostate cells, including fragments,
variants and derivatives of the nucleic acids and polypeptides. The present invention also
relates to antibodies to the polypeptides of the invention, as well as agonists and
antagonists of the polypeptides of the invention. The invention also relates to
compositions comprising the nucleic acids, polypeptides, antibodies, variants,
15 derivatives, agonists and antagonists of the invention and methods for the use of these
compositions. These uses include identifying, diagnosing, monitoring, staging, imaging
and treating prostate cancer and non-cancerous disease states in prostate tissue,
identifying prostate tissue and monitoring and identifying and/or designing agonists and
antagonists of polypeptides of the invention. The uses also include gene therapy,
20 production of transgenic animals and cells, and production of engineered prostate tissue
for treatment and research.

BACKGROUND OF THE INVENTION

Prostate cancer is the most prevalent cancer in men and is the second leading
cause of death from cancer among males in the United States. AJCC Cancer Staging
25 Handbook 203 (Irvin D. Fleming et al. eds., 5th ed. 1998); Walter J. Burdette, Cancer:
Etiology, Diagnosis, and Treatment 147 (1998). In 1999, it was estimated that 37,000
men in the United States would die as result of prostate cancer. Elizabeth A. Platz et al.,
& Edward Giovannucci, *Epidemiology of and Risk Factors for Prostate Cancer*, in
Management of Prostate Cancer 21 (Eric A Klein, ed. 2000). Cancer of the prostate
30 typically occurs in older males, with a median age of 74 years for clinical diagnosis.

Burdette, *supra* at 147. A man's risk of being diagnosed with invasive prostate cancer in his lifetime is one in six. Platz et al., *supra* at 21.

Although our understanding of the etiology of prostate cancer is incomplete, the results of extensive research in this area point to a combination of age, genetic and

5 environmental/dietary factors. Platz et al., *supra* at 19; Burdette, *supra* at 147; Steven K. Clinton, *Diet and Nutrition in Prostate Cancer Prevention and Therapy*, in Prostate Cancer: A Multidisciplinary Guide 246-269 (Philip W. Kantoff et al. eds. 1997).

Broadly speaking, genetic risk factors predisposing one to prostate cancer include race and a family history of the disease. Platz et al., *supra* at 19, 28-29, 32-34. Aside from
10 these generalities, a deeper understanding of the genetic basis of prostate cancer has remained elusive. Considerable research has been directed to studying the link between prostate cancer, androgens, and androgen regulation, as androgens play a crucial role in prostate growth and differentiation. Meena Augustus et al., *Molecular Genetics and Markers of Progression*, in Management of Prostate Cancer 59 (Eric A Klein ed. 2000).
15 While a number of studies have concluded that prostate tumor development is linked to elevated levels of circulating androgen (*e.g.*, testosterone and dihydrotestosterone), the genetic determinants of these levels remain unknown. Platz et al., *supra* at 29-30.

Several studies have explored a possible link between prostate cancer and the androgen receptor (AR) gene, the gene product of which mediates the molecular and
20 cellular effects of testosterone and dihydrotestosterone in tissues responsive to androgens. *Id.* at 30. Differences in the number of certain trinucleotide repeats in exon 1, the region involved in transactivational control, have been of particular interest. Augustus et al., *supra* at 60. For example, these studies have revealed that as the number of CAG repeats decreases the transactivation ability of the gene product increases, as
25 does the risk of prostate cancer. Platz et al., *supra* at 30-31. Other research has focused on the α -reductase Type 2 gene, the gene which codes for the enzyme that converts testosterone into dihydrotestosterone. *Id.* at 30. Dihydrotestosterone has greater affinity for the AR than testosterone, resulting in increased transactivation of genes responsive to androgens. *Id.* While studies have reported differences among the races in the length of
30 a TA dinucleotide repeat in the 3' untranslated region, no link has been established between the length of that repeat and prostate cancer. *Id.*

Interestingly, while *ras* gene mutations are implicated in numerous other cancers, such mutations appear not to play a significant role in prostate cancer, at least among Caucasian males. Augustus, *supra* at 52.

- Environmental/dietary risk factors which may increase the risk of prostate cancer include intake of saturated fat and calcium. Platz et al., *supra* at 19, 25-26. Conversely, intake of selenium, vitamin E and tomato products (which contain the carotenoid lycopene) apparently decrease that risk. *Id.* at 19, 26-28 The impact of physical activity, cigarette smoking, and alcohol consumption on prostate cancer is unclear. Platz et al., *supra* at 23-25.
- Periodic screening for prostate cancer is most effectively performed by digital rectal examination (DRE) of the prostate, in conjunction with determination of the serum level of prostate-specific antigen (PSA). Burdette, *supra* at 148. While the merits of such screening are the subject of considerable debate, Jerome P. Richie & Irving D. Kaplan, *Screening for Prostate Cancer: The Horns of a Dilemma*, in Prostate Cancer: A Multidisciplinary Guide 1-10 (Philip W. Kantoff et al. eds. 1997), the American Cancer Society and American Urological Association recommend that both of these tests be performed annually on men 50 years or older with a life expectancy of at least 10 years, and younger men at high risk for prostate cancer. Ian M. Thompson & John Foley, *Screening for Prostate Cancer*, in Management of Prostate Cancer 71 (Eric A Klein ed. 2000). If necessary, these screening methods may be followed by additional tests, including biopsy, ultrasonic imaging, computerized tomography, and magnetic resonance imaging. Christopher A. Haas & Martin I. Resnick, *Trends in Diagnosis, Biopsy, and Imaging*, in Management of Prostate Cancer 89-98 (Eric A Klein ed. 2000); Burdette, *supra* at 148.
- Once the diagnosis of prostate cancer has been made, treatment decisions for the individual are typically linked to the stage of prostate cancer present in that individual, as well as his age and overall health. Burdette, *supra* at 151. One preferred classification system for staging prostate cancer was developed by the American Urological Association (AUA). *Id.* at 148. The AUA classification system divides prostate tumors into four broad stages, A to D, which are in turn accompanied by a number of smaller substages. Burdette, *supra* at 152-153; Anthony V. D'Amico et al., *The Staging of*

Prostate Cancer, in Prostate Cancer: A Multidisciplinary Guide 41 (Philip W. Kantoff et al. eds. 1997).

Stage A prostate cancer refers to the presence of microscopic cancer within the prostate gland. D'Amico, *supra* at 41. This stage is comprised of two substages: A1, which involves less than four well-differentiated cancer foci within the prostate, and A2, which involves greater than three well-differentiated cancer foci or alternatively, moderately to poorly differentiated foci within the prostate. Burdette, *supra* at 152; D'Amico, *supra* at 41. Treatment for stage A1 preferentially involves following PSA levels and periodic DRE. Burdette, *supra* at 151. Should PSA levels rise, preferred treatments include radical prostatectomy in patients 70 years of age and younger, external beam radiotherapy for patients between 70 and 80 years of age, and hormone therapy for those over 80 years of age. *Id.*

Stage B prostate cancer is characterized by the presence of a palpable lump within the prostate. Burdette, *supra* at 152-53; D'Amico, *supra* at 41. This stage is comprised of three substages: B1, in which the lump is less than 2 cm and is contained in one lobe of the prostate; B2, in which the lump is greater than 2 cm yet is still contained within one lobe; and B3, in which the lump has spread to both lobes. Burdette, *supra*, at 152-53. For stages B1 and B2, the treatment again involves radical prostatectomy in patients 70 years of age and younger, external beam radiotherapy for patients between 70 and 80 years of age, and hormone therapy for those over 80 years of age. *Id.* at 151. In stage B3, radical prostatectomy is employed if the cancer is well-differentiated and PSA levels are below 15 ng/mL; otherwise, external beam radiation is the chosen treatment option. *Id.*

Stage C prostate cancer involves a substantial cancer mass accompanied by extraprostatic extension. Burdette, *supra* at 153; D'Amico, *supra* at 41. Like stage A prostate cancer, Stage C is comprised of two substages: substage C1, in which the tumor is relatively minimal, with minor prostatic extension, and substage C2, in which the tumor is large and bulky, with major prostatic extension. *Id.* The treatment of choice for both substages is external beam radiation. Burdette, *supra* at 151.

The fourth and final stage of prostate cancer, Stage D, describes the extent to which the cancer has metastasized. Burdette, *supra* at 153; D'Amico, *supra* at 41. This stage is comprised of four substages: (1) D0, in which acid phosphatase levels are

persistently high, (2) D1, in which only the pelvic lymph nodes have been invaded, (3) D2, in which the lymph nodes above the aortic bifurcation have been invaded, with or without distant metastasis, and (4) D3, in which the metastasis progresses despite intense hormonal therapy. *Id.* Treatment at this stage may involve hormonal therapy, chemotherapy, and removal of one or both testes. Burdette, *supra* at 151.

Despite the need for accurate staging of prostate cancer, current staging methodology is limited. The wide variety of biological behavior displayed by neoplasms of the prostate has resulted in considerable difficulty in predicting and assessing the course of prostate cancer. Augustus et al., *supra* at 47. Indeed, despite the fact that most prostate cancer patients have carcinomas that are of intermediate grade and stage, prognosis for these types of carcinomas is highly variable. Andrew A Renshaw & Christopher L. Corless, *Prognostic Features in the Pathology of Prostate Cancer*, in Prostate Cancer: A Multidisciplinary Guide 26 (Philip W. Kantoff et al. eds. 1997). Techniques such as transrectal ultrasound, abdominal and pelvic computerized tomography, and MRI have not been particularly useful in predicting local tumor extension. D'Amico, *supra* at 53 (editors' comment). While the use of serum PSA in combination with the Gleason score is currently the most effective method of staging prostate cancer, *id.*, PSA is of limited predictive value, Augustus et al., *supra* at 47; Renshaw et al., *supra* at 26, and the Gleason score is prone to variability and error, King, C. R. & Long, J. P., *Int'l. J. Cancer* 90(6): 326-30 (2000). As such, the current focus of prostate cancer research has been to obtain biomarkers to help better assess the progression of the disease. Augustus et al., *supra* at 47; Renshaw et al., *supra* at 26; Pettaway, C. A., *Tech. Urol.* 4(1): 35-42 (1998).

Accordingly, there is a great need for more sensitive and accurate methods for predicting whether a person is likely to develop prostate cancer, for diagnosing prostate cancer, for monitoring the progression of the disease, for staging the prostate cancer, for determining whether the prostate cancer has metastasized and for imaging the prostate cancer. There is also a need for better treatment of prostate cancer.

30

SUMMARY OF THE INVENTION

The present invention solves these and other needs in the art by providing nucleic acid molecules and polypeptides as well as antibodies, agonists and antagonists, thereto

that may be used to identify, diagnose, monitor, stage, image and treat prostate cancer and non-cancerous disease states in prostate; identify and monitor prostate tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and
5 cells, and methods for producing engineered prostate tissue for treatment and research.

Accordingly, one object of the invention is to provide nucleic acid molecules that are specific to prostate cells and/or prostate tissue. These prostate specific nucleic acids (PSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If
10 the PSNA is genomic DNA, then the PSNA is a prostate specific gene (PSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to prostate. In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 137 through 240. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid
15 sequence of SEQ ID NO: 1 through 136. By nucleic acid molecule, it is also meant to be inclusive of sequences that selectively hybridize or exhibit substantial sequence similarity to a nucleic acid molecule encoding a PSP, or that selectively hybridize or exhibit substantial sequence similarity to a PSNA, as well as allelic variants of a nucleic acid molecule encoding a PSP, and allelic variants of a PSNA. Nucleic acid molecules
20 comprising a part of a nucleic acid sequence that encodes a PSP or that comprises a part of a nucleic acid sequence of a PSNA are also provided.

A related object of the present invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of a PSNA. In a preferred embodiment, the nucleic acid
25 molecule comprises one or more expression control sequences controlling the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of a PSP.

Another object of the invention is to provide vectors and/or host cells comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic
30 acid molecule encodes all or a fragment of a PSP. In another preferred embodiment, the nucleic acid molecule comprises all or a part of a PSNA.

Another object of the invention is to provided methods for using the vectors and host cells comprising a nucleic acid molecule of the instant invention to recombinantly produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the invention. In a preferred embodiment, the polypeptide is a PSP. The polypeptide may comprise either a fragment or a full-length protein as well as a mutant protein (mutein), fusion protein, homologous protein or a polypeptide encoded by an allelic variant of a PSP.

Another object of the invention is to provide an antibody that specifically binds to a polypeptide of the instant invention..

Another object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

Another object of the invention is to provide methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. In a preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and treating prostate cancer and non-cancerous disease states in prostate. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying and/or monitoring prostate tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered prostate tissue for treatment and research.

The polypeptides and/or antibodies of the instant invention may also be used to identify, diagnose, monitor, stage, image and treat prostate cancer and non-cancerous disease states in prostate. The invention provides methods of using the polypeptides of the invention to identify and/or monitor prostate tissue, and to produce engineered prostate tissue.

The agonists and antagonists of the instant invention may be used to treat prostate cancer and non-cancerous disease states in prostate and to produce engineered prostate tissue.

Yet another object of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences.

5 DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular
10 terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques of the present invention are generally performed
15 according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook et al., Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor
20 Press (2001); Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel et al., Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using
25 Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999); each of which is incorporated herein by reference in its entirety.*

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and
30 techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in

the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

- 5 A “nucleic acid molecule” of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A “nucleic acid molecule” as used herein is synonymous with “nucleic acid” and
- 10 “polynucleotide.” The term “nucleic acid molecule” usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.
- 15 The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates,
- 20 phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term “nucleic acid molecule” also includes any topological conformation, including single-stranded, double-stranded, partially
- 25 duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.
- 30 A “gene” is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may

comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a
5 nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute a contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

10 A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an
15 RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a
20 portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized
25 polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

30 A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid

molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g., Geysen et al., Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are

not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized

5 oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

10 The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate, 15 phosphorodiselenoate, phosphoroanilothioate, phosphoraniladate, phosphoroamidate, and the like. See e.g., LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); United States Patent No. 20 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is 25 referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of 30 RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence.

In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term “percent sequence identity” in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms “percent sequence identity”, “percent sequence similarity” and “percent sequence homology” interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% sequence identity, over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. See Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$$T_m = 81.5^{\circ}\text{C} + 16.6 (\log_{10}[\text{Na}^+]) + 0.41 (\text{fraction G} + \text{C}) - 0.63 (\% \text{ formamide}) - (600/l)$$

where l is the length of the hybrid in base pairs.

The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

5 $T_m = 79.8^{\circ}\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.35$
(% formamide) - (820/ l).

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$$T_m = 79.8^{\circ}\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) + 11.8 (\text{fraction G} + \text{C})^2 - 0.50$$

(% formamide) - (820/ l).

10 In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C
15 would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

20 An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without
25 formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a
30 library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping

the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (e.g. 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well-known in the art. *See* Sambrook *et al.*

- 5 (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

- 15 As defined herein, nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid molecule is created synthetically or recombinantly using high codon degeneracy as permitted by the redundancy of the genetic code.

- 20 Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (e.g., for oligonucleotide probes) may be calculated by the formula: $T_m = 81.5^{\circ}\text{C} + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\text{fraction G+C}) - (600/N)$, wherein N is change length and the $[\text{Na}^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or “guessmers,” as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. *See, e.g.,* Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-30 11.57.

The term “digestion” or “digestion of DNA” refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The

various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 µg of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 µl of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 µg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAs. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity

with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Microarrays or nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).

The term "mutated" when applied to nucleic acid molecules means that nucleotides in the nucleic acid sequence of the nucleic acid molecule may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the nucleic acid molecule comprises the wild type nucleic acid sequence encoding a PSP or is a PSNA. The nucleic acid molecule may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g.*, Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR

reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

The term “sexual PCR mutagenesis” or “DNA shuffling” refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g., Stemmer, Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes (“Family shuffling”).

10 The term “*in vivo* mutagenesis” refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These “mutator” strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term “cassette mutagenesis” refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide “cassette” that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

20 The term “recursive ensemble mutagenesis” refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. *See, e.g., Arkin et al., Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

25 The term “exponential ensemble mutagenesis” refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. *See, e.g., Delegrave et al., Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993). Each of the references mentioned above are hereby incorporated by reference in its entirety.

“Operatively linked” expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

5 The term “expression control sequence” as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination,
10 promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such
15 control sequences generally include the promoter, ribosomal binding site, and transcription termination sequence. The term “control sequences” is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

20 The term “vector,” as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a “plasmid”, which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of
25 vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and
30 thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as “recombinant expression vectors” (or simply, “expression

vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector.

However, the invention is intended to include other forms of expression vectors that
5 serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which an expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding
10 generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of a transcript-derived nucleic acid that can be translated in
15 its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

20 As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally-occurring and non-naturally-
25 occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises a PSP encoded by a nucleic acid molecule of the instant
30 invention, as well as a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally

associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be

5 "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single

10 species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample,

15 followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "polypeptide fragment" as used herein refers to a polypeptide of the instant invention that has an amino-terminal and/or carboxy-terminal deletion compared

20 to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40

25 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such

30 modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid

derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, 5 proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of 10 substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H , ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, 15 stability requirements, and available instrumentation. Methods for labeling polypeptides are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

The term "fusion protein" refers to polypeptides of the instant invention comprising polypeptides or fragments coupled to heterologous amino acid sequences. 20 Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be 25 produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

30 The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide of the instant invention that is comprised of a segment of at least 25 amino acids that has substantial

identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide of the instant invention. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH₂NH--, --CH₂S--, --CH₂-CH₂--, --CH=CH--(cis and trans), --COCH₂--, --CH(OH)CH₂--, and --CH₂SO--, by methods well-known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al.*, *Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

A "polypeptide mutant" or "mutein" refers to a polypeptide of the instant invention whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid

- sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.
- Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991), each of which are incorporated herein by reference.

- As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. See Golub *et al.* (eds.), Immunology - A Synthesis 2nd Ed., Sinauer Associates (1991), which is incorporated herein by reference. Stereoisomers (*e.g.*, D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α -, β -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino

acids may also be suitable components for polypeptides of the present invention.

Examples of unconventional amino acids include: 4-hydroxyproline, γ -carboxyglutamate, -N,N,N-trimethyllysine, -N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other
5 similar amino acids and imino acids (e.g., 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

A protein has "homology" or is "homologous" to a protein from another organism
10 if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological activity or function. Alternatively, a protein may have homology or be homologous to another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous,"
15 this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous protein is one that exhibits 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence
20 similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino
25 acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In general, a conservative amino
30 acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted

upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. *See, e.g., Pearson, Methods Mol. Biol.* 24: 307-31 (1994), herein incorporated by reference.

For instance, the following six groups each contain amino acids that are

5 conservative substitutions for one another:

- 1) Serine (S), Threonine (T);
- 2) Aspartic Acid (D), Glutamic Acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 10 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992), herein incorporated by reference. A “moderately conservative” replacement is
 15 any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid
 20 substitutions. For instance, GCG contains programs such as “Gap” and “Bestfit” which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. *See, e.g., GCG Version 6.1. Other programs include FASTA, discussed supra.*

25 A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. *See, e.g., Altschul et al., J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997); herein incorporated by reference. Preferred parameters for blastp are:

- 30 Expectation value: 10 (default)
 Filter: seg (default)
 Cost to open a gap: 11 (default)

-28-

- Cost to extend a gap: 1 (default)
- Max. alignments: 100 (default)
- Word size: 11 (default)
- No. of descriptions: 100 (default)
- 5 Penalty Matrix: BLOSUM62

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number
10 of different organisms, it is preferable to compare amino acid sequences.

Database searching using amino acid sequences can be measured by algorithms other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best
15 overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

20 An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', F(ab')₂, Fv,
25 dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an F(ab')₂ fragment is a bivalent fragment comprising two Fab
30 fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single

arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g., Ward et al., Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which a VL and VH region are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g., Bird et al., Science* 242: 423-426 (1988); Huston et al., *Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g., Holliger et al., Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak et al., *Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that

purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (e.g., BSA) or a chemical such as polyethylene glycol (PEG).

A “neutralizing antibody” or “an inhibitory antibody” is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An “activating antibody” is an antibody that increases the activity of a polypeptide.

The term “epitope” includes any protein determinant capable of specifically binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than $1\ \mu\text{M}$, preferably less than $100\ \text{nM}$ and most preferably less than $10\ \text{nM}$.

The term “patient” as used herein includes human and veterinary subjects.

Throughout this specification and claims, the word “comprise,” or variations such as “comprises” or “comprising,” will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term “prostate specific” refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the prostate as compared to other tissues in the body. In a preferred embodiment, a “prostate specific” nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the “prostate specific” nucleic acid molecule or polypeptide is expressed at a level that is 10-fold higher than any other tissue in the body, more preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels, such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

5 One aspect of the invention provides isolated nucleic acid molecules that are specific to the prostate or to prostate cells or tissue or that are derived from such nucleic acid molecules. These isolated prostate specific nucleic acids (PSNAs) may comprise a cDNA, a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to prostate, a prostate-specific polypeptide (PSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 137 through 240. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 136.

15 A PSNA may be derived from a human or from another animal. In a preferred embodiment, the PSNA is derived from a human or other mammal. In a more preferred embodiment, the PSNA is derived from a human or other primate. In an even more preferred embodiment, the PSNA is derived from a human.

20 By "nucleic acid molecule" for purposes of the present invention, it is also meant to be inclusive of nucleic acid sequences that selectively hybridize to a nucleic acid molecule encoding a PSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode a PSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes a PSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 137 through 240. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1 through 136.

30 In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a PSP under low stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a PSP under moderate stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule

encoding a PSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 137 through 240. In a yet more preferred
5 embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1 through 136. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

10 By "nucleic acid molecule" as used herein it is also meant to be inclusive of sequences that exhibits substantial sequence similarity to a nucleic acid encoding a PSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human PSP. In a more preferred embodiment, the nucleic acid molecule
15 exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 137 through 240. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding a PSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 137 through 240, more preferably at least
20 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding a PSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid
25 molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding a PSP.

In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a PSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule
30 comprising a nucleic acid sequence of SEQ ID NO: 1 through 136. In a preferred embodiment, the nucleic acid molecule is one that has at least 60% sequence identity with a PSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1 through 136,

more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with a PSNA, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a PSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one that exhibits sequence identity over its entire length to a PSNA or to a nucleic acid molecule encoding a PSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the PSNA or the nucleic acid molecule encoding a PSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 137 through 240 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO: 1 through 136. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the PSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed mutation of a PSNA. Further, the substantially

similar nucleic acid molecule may or may not be a PSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is a PSNA.

By "nucleic acid molecule" it is also meant to be inclusive of allelic variants of a PSNA or a nucleic acid encoding a PSP. For instance, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes. In fact, more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001). Thus, the sequence determined from one individual of a species may differ from other allelic forms present within the population. Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the nucleic acid molecule comprising an allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes a PSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes a PSP comprising an amino acid sequence of SEQ ID NO: 137 through 240. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is a PSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1 through 136. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

By "nucleic acid molecule" it is also meant to be inclusive of a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is a PSP. However, in a preferred embodiment, the part encodes a PSP. In one aspect, the invention comprises a part of a PSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to a PSNA. In a third aspect, the invention comprises a part of a nucleic acid molecule that is an allelic variant of a PSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes a PSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides.

-35-

The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

By "nucleic acid molecule" it is also meant to be inclusive of sequence that encoding a fusion protein, a homologous protein, a polypeptide fragment, a mutein or a polypeptide analog, as described below.

Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

In a preferred embodiment of the invention, the nucleic acid molecule contains modifications of the native nucleic acid molecule. These modifications include nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

In a preferred embodiment, isolated nucleic acid molecules can include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment, the labeled nucleic acid molecule may be used as a hybridization probe.

Common radiolabeled analogues include those labeled with ³³P, ³²P, and ³⁵S, such as ³²P-dATP, ³²P-dCTP, ³²P-dGTP, ³²P-dTTP, ³²P-3'dATP, ³²P-ATP, ³²P-CTP, ³²P-GTP, ³²P-UTP, ³⁵S-dATP, α -³⁵S-GTP, α -³³P-dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. See Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids.

For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA to provide a stable coordination complex between the nucleic acid and fluorophore label
5 (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); *see Alers et al., Genes, Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al., J. NIH Res.* 5: 82 (1994); Van Belkum *et al., BioTechniques* 16: 148-153 (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker
10 (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally-coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the
15 nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. *See, e.g., Tyagi et al., Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al., Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al., Proc. Natl. Acad. Sci.*
20 *USA* 95: 11538-11543 (1998); Kostrikis *et al., Science* 279: 1228-1229 (1998); Marras *et al., Genet. Anal.* 14: 151-156 (1999); U. S. Patent 5,846,726; 5,925,517; 5,925,517; 5,723,591 and 5,538,848; Holland *et al., Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al., Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al., Nucleic Acids Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by
25 reference in their entireties.

Nucleic acid molecules of the invention may be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. *See Hartmann et al. (eds.), Manual of Antisense Methodology: Perspectives in*
30 *Antisense Science*, Kluwer Law International (1999); Stein *et al. (eds.), Applied Antisense Oligonucleotide Technology*, Wiley-Liss (1998); Chadwick *et al. (eds.), Oligonucleotides as Therapeutic Agents - Symposium No. 209*, John Wiley & Son Ltd (1997); the disclosures of which are incorporated herein by reference in their entireties.

Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction. See Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation,

- 5 phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having
- 10 normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U. S. Patents 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019;
- 15 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

- 20 Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a
- 25 nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts. Representative U.S. patents
- 30 that teach the preparation of the above backbones include, but are not limited to, U.S. Patent 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307;

5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Patent 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (see, e.g., "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The T_m of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the T_m of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the T_m by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the T_m by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyamide backbone with nucleobase sidechains. See, e.g., Ray *et al.*, *FASEB J.* 14(9): 1041–60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1):

3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta*. 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules may be modified compared to their native structure
5 throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in U.S. Patents 5,760,012 and 5,731,181, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24:
10 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed,
15 partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by
20 reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta*. 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

25

Methods for Using Nucleic Acid Molecules as Probes and Primers

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic
30 acid samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of a PSNA, such as deletions, insertions, translocations, and duplications of the PSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g., Andreeff et al. (eds.), Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications*, John Wiley & Sons (1999), the disclosure of which is incorporated herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include the nucleic acid molecules of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect, characterize, and quantify PSNA in, and isolate PSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A⁺- selected RNA samples. In another aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g., Schwarchzacher et al., In Situ Hybridization*, Springer-Verlag New York (2000), the disclosure of which is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to PSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*;

Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in their entirety.

- Thus, in one embodiment, a nucleic acid molecule of the invention may be used
- 5 as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a PSP. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 137 through 240.
- 10 In another preferred embodiment, the probe or primer is derived from a PSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 136.

- In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17
- 15 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using
- 20 oligonucleotide probes are well-known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

- Methods of performing primer-directed amplification are also well-known in the
- 25 art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques,
- 30 John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson, *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of

which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique:RT-PCR, Eaton Publishing Company/ BioTechniques Books
5 (1995); the disclosure of which is incorporated herein by reference in its entirety.

PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion
10 protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Patent 6,004,744,
15 the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); U.S. Patents 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by
20 reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-
25 planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or
30 positively-charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another

embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention relates to vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acids of the present invention, alone or as fusions to heterologous

polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

Vectors are by now well-known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are
10 incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that
15 expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or
20 GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

25 In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial
30 hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous

derivatives of phage lambda, *e.g.*, NM989, λ GT10 and λ GT11, and other phages, *e.g.*, M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEp series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2 μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as *ura3-52*, *his3-D1*, *leu2-D1*, *trp1-D1* and *lys2-201*.

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda*, *e.g.*, Sf9 and Sf21 cell lines, and expresSFTM cells (Protein Sciences Corp., Meriden, CT, USA)), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between

these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

In another embodiment, the host cells may be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention may be modified to resemble, as much as possible, genes naturally

contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the *trc* promoter, a hybrid derived from the *trp* and *lac* promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of *fd* coat protein, or the *araBAD* operon. Prokaryotic expression vectors may further include transcription terminators, such as the *aspA* terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the *CYC1* promoter, the *GAL1* promoter, the *GAL10* promoter, *ADH1* promoter, the promoters of the yeast α -mating system, or the *GPD* promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the *CYC1* or *ADH1* gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the

Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the PSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β -globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the *trc* promoter, which is regulated by the *lac* operon, and the *pL* promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid *Plac/ara-1* promoter and the *PLtetO-1* promoter. The *PLtetO-1* promoter takes advantage of the high expression levels from the *PL* promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the *Tn10* tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline

(Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverly, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion protein with glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that

carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusion to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusion proteins for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the α -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from

those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the art. *See* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g.* Cormack *et al.*, *Gene* 173: 33-38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g.* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.* Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patents 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. *See also* Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application Nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable

expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA), allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation,

and acylation, and it is an aspect of the present invention to provide PSPs with such post-translational modifications.

Polypeptides of the invention may be post-translationally modified. Post-translational modifications include phosphorylation of amino acid residues serine, threonine and/or tyrosine, N-linked and/or O-linked glycosylation, methylation, acetylation, prenylation, methylation, acetylation, arginylation, ubiquination and racemization. One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., www.expasy.org (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications may be found in web sites such as the Delta Mass database <http://www.abrf.org/ABRF/Research/Committees/deltamass/deltamass.html> (accessed October 19, 2001); "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" Cooper et al. *Nucleic Acids Res.* 29; 332-335 (2001) and <http://www.glycosuite.com/> (accessed October 19, 2001); "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. *Nucleic Acids Research*, 27: 370-372 (1999) and <http://www.cbs.dtu.dk/databases/OGLYCBASE/> (accessed October 19, 2001); "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. *Nucleic Acids Res* 27(1):237-239 (1999) and <http://www.cbs.dtu.dk/databases/PhosphoBase/> (accessed October 19, 2001); or <http://pir.georgetown.edu/pirwww/search/textresid.html> (accessed October 19, 2001).

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signaling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur

in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g., p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the

desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one
5 may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule is then be introduced into a host cell that is
10 capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

15 In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the
20 product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

The recombinant nucleic acid molecules and more particularly, the expression
25 vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid sequences according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have
30 biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors

typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell
5 (e.g., conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (*See, for instance, Ausubel, supra, and Sambrook et al., supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an
10 expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be
15 able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera*
20 *frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from
25 *Spodoptera frugiperda*, e.g., SF9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3
30 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (e.g., from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and

readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from prostate are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human prostate cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. *See, e.g.,* Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (*e.g.,* Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, *e.g.,* with CaCl_2 , or a solution of Mg^{2+} , Mn^{2+} , Ca^{2+} , Rb^+ or K^+ , dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (*e.g.,* Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5 competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent *E. coli* Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent, that is, competent to take up exogenous DNA by electroporation, by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided online in [Electroprotocols](#)

(BioRad, Richmond, CA, USA) (http://www.biorad.com/LifeScience/pdf/New_Gene_Pulser.pdf).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the
5 action of hydrolytic enzymes such as snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus*, to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+} . Subsequently, the cells are resuspended in a solution of sorbitol, mixed
10 with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective
15 medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension
20 pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

25 Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with CaPO_4 or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for CaPO_4 transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated
30 transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent,

FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA).

Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) (<http://www.bio-rad.com/LifeScience/pdf/>

- 5 New_Gene_Pulser.pdf); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.,* Cheng *et al.*, *Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA*
10 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

- Purification of recombinantly expressed proteins is now well by those skilled in the art. *See, e.g.,* Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid
15 Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.),
20 Protein Purification Applications, Oxford University Press (2001); the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

- Briefly, however, if purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by
25 means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

Polypeptides

- 30 Another object of the invention is to provide polypeptides encoded by the nucleic acid molecules of the instant invention. In a preferred embodiment, the polypeptide is a prostate specific polypeptide (PSP). In an even more preferred embodiment, the

polypeptide is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 137 through 240. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide, wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide fragment is a fragment of a PSP. In a more preferred embodiment, the fragment is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 137 through 240. A polypeptide that comprises only a fragment of an entire PSP may or may not be a polypeptide that is also a PSP. For instance, a full-length polypeptide may be prostate-specific, while a fragment thereof may be found in other tissues as well as in prostate. A polypeptide that is not a PSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-PSP antibodies. However, in a preferred embodiment, the part or fragment is a PSP. Methods of determining whether a polypeptide is a PSP are described *infra*.

Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g., Geysen et al., Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and U.S. Patents 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. *See, e.g., Lerner, Nature* 299: 592-596 (1982); Shinnick *et al., Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al., Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic, meaning that they are capable of eliciting antibody for

the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Patents 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

10 The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger fragments having at least 75
15 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments of a polypeptide by truncating the nucleic acid molecule, *e.g.*, a PSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically
20 synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. *See, e.g.*, Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a
25 fragment of polypeptide of the invention, preferably a PSP, may be produced by chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of the polypeptide, preferably a PSP, in a host cell.

By "polypeptides" as used herein it is also meant to be inclusive of mutants,
30 fusion proteins, homologous proteins and allelic variants of the polypeptides specifically exemplified.

A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be prostate-specific. In a preferred embodiment, the mutein is prostate-specific. In a preferred embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 137 through 240. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO: 137 through 240. In yet a more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO: 137 through 240.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein may be produced from a host cell comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is prostate-specific, as described below. Multiple random mutations can be introduced into the gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo*

mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.,* Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel
5 (1999), U.S. Patent 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

By "polypeptide" as used herein it is also meant to be inclusive of polypeptides homologous to those polypeptides exemplified herein. In a preferred embodiment, the polypeptide is homologous to a PSP. In an even more preferred embodiment, the
10 polypeptide is homologous to a PSP selected from the group having an amino acid sequence of SEQ ID NO: 137 through 240. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to a PSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity to an comprising an amino acid sequence of SEQ ID NO: 137 through 240. In an even
15 more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO: 137 through 240. In a yet more preferred embodiment, the homologous polypeptide is one that exhibits at least 85%,
20 more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to a PSP comprising an amino acid sequence of SEQ ID NO: 137 through 240. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to a PSP comprising an amino acid sequence
25 of SEQ ID NO: 137 through 240. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to a PSNA. In a preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a
30 PSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the PSNA is selected from the group consisting of SEQ ID NO: 1 through 136. In another preferred embodiment, the

homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes a PSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the PSP is selected from the group consisting of SEQ ID NO: 137 through 240.

- 5 The homologous polypeptide may be a naturally-occurring one that is derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 137 through 240. The homologous polypeptide may also be a naturally-
10 occurring polypeptide from a human, when the PSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring
15 polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. In another embodiment, the homologous polypeptide may be
20 one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of a PSP. Further, the homologous protein may or may not encode polypeptide that is a PSP. However, in a
25 preferred embodiment, the homologous polypeptide encodes a polypeptide that is a PSP.

- Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but
30 also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the

present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, by "polypeptide" as used herein it is also meant to be inclusive of polypeptides encoded by an allelic variant of a nucleic acid molecule encoding a PSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 137 through 240. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 136.

In another embodiment, the invention provides polypeptides which comprise derivatives of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a PSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 137 through 240, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H . In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983);

Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591,

BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

- 5 The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS,
- 10 HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC,
- 15 SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

- The polypeptides, fragments, and fusion proteins of the present invention can be
- 20 conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

- The polypeptides, fragments, and fusion proteins of the present invention can also
- 25 usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-PSP antibodies.

- The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum
- 30 half-life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999),

incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

- 5 In yet another embodiment, the invention provides analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a PSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 137 through 240. In a preferred embodiment, the analog is one that comprises one or
- 10 more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In general, the non-peptide analog is structurally similar to a PSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH₂NH--, --CH₂S--, --CH₂-CH₂--,
- 15 --CH=CH--(cis and trans), --COCH₂--, --CH(OH)CH₂-- and --CH₂SO--. In another embodiment, the non-peptide analog comprises substitution of one or more amino acids of a PSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific
- 20 three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g., Kole et al., Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.
- 25 Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000);
- 30 Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer

Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entireties.

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinoyl-(9-fluorenylmethoxycarbonyl)-L-lysine (Fmoc biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The Fmoc and *t*BOC derivatives of dabcyL-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyL chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyL quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS-Fmoc-L-glutamic acid or the corresponding *t*BOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated Fmoc synthesis of peptides using (Fmoc)-TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other Fmoc-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-

2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)- β -alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

Fusion Proteins

The present invention further provides fusions of each of the polypeptides and fragments of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide is a PSP. In a more preferred embodiment, the polypeptide that is fused to the heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 137 through 240, or is a mutein, homologous polypeptide;

analog or derivative thereof. In an even more preferred embodiment, the nucleic acid molecule encoding the fusion protein comprises all or part of the nucleic acid sequence of SEQ ID NO: 1 through 136, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 136.

The fusion proteins of the present invention will include at least one fragment of the protein of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the protein of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particularly useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g., Ausubel, Chapter 16, (1992), supra.* Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation

- of secretion signals and/or leader sequences. For example, a His⁶ tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.
- 10 Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. See Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595; Fabbri *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A.* 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-
25 hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in *Drosophila* with peptide aptamers. *Proc Natl Acad Sci U S A* 95, 14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein
30 interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623-627; Ito, *et al.*, (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by

reference in their entireties. Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded
5 protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

The polypeptides and fragments of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin
10 A, *anthrax* toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, myc, hemagglutinin (HA), GST, immunoglobulins, β -galactosidase, biotin trpE, protein A, β -lactamase, -amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine
15 at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast _ mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. *See, e.g.*, Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by
20 enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (*e.g.*, a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind
25 the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the PSP.

As further described below, the isolated polypeptides, muteins, fusion proteins, homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize PSPs, their allelic
30 variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly PSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser

scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of PSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of PSPs.

- 5 One may determine whether polypeptides including muteins, fusion proteins, homologous proteins or allelic variants are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon
10 linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-
15 102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides including fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of one having ordinary skill in the art. *See, e.g.*, Scopes, Protein Purification, 2d ed. (1987).

- 20 Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

- Accordingly, it is an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-
25 proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

- Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents, such as in vaccines and as replacement therapy, the isolated proteins of the present invention are also useful at lower purity. For
30 example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction there between.

Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies are specific for a polypeptide that is a PSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 137 through 240, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a PSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or *visa versa*. In addition, alternative splice forms of a PSP may be indicative of cancer. Differential degradation of the C or N-terminus of a PSP may also be a marker or target for anticancer therapy. For example, a PSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is well-known in the art, the degree to which an antibody can discriminate among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-PSP polypeptides by at least 2-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the protein of the present invention in samples derived from human prostate.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the

present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7} M, 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and up to 1×10^{-13} M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as
5 IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In this case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the protein or protein fragments of the present invention. Such
10 antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic
15 mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patents 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by
20 reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered
25 antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention can also be obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster) lagomorphs, typically rabbits, and also larger
30 mammals, such as sheep, goats, cows, and horses, and other egg laying birds or reptiles such as chickens or alligators. For example, avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000, the contents of which are

hereby incorporated in their entirety. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein or protein fragment of the present invention.

5 As discussed above, virtually all fragments of 8 or more contiguous amino acids of the proteins of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

10 Immunogenicity can also be conferred by fusion of the polypeptide and fragments of the present invention to other moieties. For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725
15 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. See Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996), the disclosures of which are incorporated herein by reference. Immunization protocols often include multiple immunizations, either with or
20 without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990)).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the proteins of the present invention and monoclonal
25 antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Antibodies from avian species may have particular
30

advantage in detection of the proteins of the present invention, in human serum or tissues (Vikinge et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998)).

Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, 5 *supra*; Zola, *supra*; Howard et al. (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their 10 entireties, and thus need not be detailed here.

Briefly, however, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding 15 antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the proteins and protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S. Patent 20 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant production of either whole antibodies, antibody 25 fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein 30 (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.*, Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths et al., *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom et al., *Immunotechnology*,

4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997); Aujame *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994).

- 5 Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.*, Barbas (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entirety.

- Typically, phage-displayed antibody fragments are scFv fragments or Fab
10 fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

- 15 For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.*, Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, *J. Biotechnol.* 76(2-3):1 57-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.*
20 201(1): 67-75 (1997); Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein by reference in their entirety.

- Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.*, Li *et al.*, *Protein Expr. Purif.*
25 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997); and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entirety.

- Antibodies and fragments and derivatives thereof of the present invention can
30 also be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gaviolondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*,

Biotechnol. Appl. Biochem. 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein by reference in their entireties.

5 Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock *et al.*, *J. Immunol. Methods.* 231: 147-57 (1999); Young *et al.*, *Res. Immunol.* 149: 609-10 (1998); Limonta *et al.*, *Immunotechnology* 1: 107-13 (1995), the disclosures of which are incorporated herein by reference in their entireties.

10 Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated by reference, review and compare bacterial, yeast, insect and mammalian expression
15 systems for expression of antibodies.

Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem. (Tokyo)* 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57
20 (1999), the disclosures of which are incorporated herein by reference in their entireties.

The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the
25 proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful fragments are Fab, Fab', Fv, F(ab)'₂, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

30 It is also an aspect of the present invention to provide antibody derivatives that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated

nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

5 Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half life of the antibodies.

10 Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.,* United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA* 81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature*
15 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entireties. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2
20 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties.

Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain
25 diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention includes any recombinant vector containing the coding sequences, or part thereof,
30 whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions

including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label is preferably an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase, β -galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopyranoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate

reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. See, e.g., Thorpe *et al.*, *Methods Enzymol.* 133: 331-53 (1986); Kricka *et al.*, *J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al.*, *J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

10 As another example, when the antibodies of the present invention are used, e.g., for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

15 For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

20 Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568,
25 BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5,
30 Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for Western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P , ^{32}P , ^{35}S , ^3H ,
5 and ^{125}I .

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th , ^{227}Ac , ^{225}Ac , ^{223}Ra , ^{213}Bi , ^{212}Pb , ^{212}Bi , ^{211}At , ^{203}Pb , ^{194}Os , ^{188}Re , ^{186}Re , ^{153}Sm , ^{149}Tb , ^{131}I , ^{125}I , ^{111}In , ^{105}Rh , $^{99\text{m}}\text{Tc}$, ^{97}Ru , ^{90}Y , ^{90}Sr , ^{88}Y , ^{72}Se , ^{67}Cu , or ^{47}Sc .

10 As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the
15 application for which they are mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, diphtheria
20 toxin, shiga toxin A, anthrax toxin lethal factor, or ricin. See Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

The antibodies of the present invention can usefully be attached to a substrate,
25 and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more
30 of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to
5 paramagnetic microspheres, typically by biotin-streptavidin interaction, which microspheres can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in
10 prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind
15 specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present
20 invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody
25 molecule, or to alter it in any other way that may render it more suitable for a particular application.

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human
30 organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding a PSP. In a preferred embodiment, the PSP comprises an amino acid

sequence selected from SEQ ID NO: 137 through 240, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise a PSNA of the invention, preferably a PSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 136, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human PSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g.,* Hogan *et al.*, Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g.,* Paterson *et al.*, *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11: 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and U.S. Patent 4,873,191 (1989) retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g.,* Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g.,* Thompson *et al.*, *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g.,* Lo, 1983, *Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g.,* Ulmer *et al.*, *Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g.,* Lavitrano *et al.*, *Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g.,* Campell *et al.*, *Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e., a*

nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene
5 may also be selectively introduced into and activated in a particular cell type by following, e.g., the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

10 Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using
15 techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

20 Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels
25 because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is
30 appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of

the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g., Gu et al., Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g., Smithies et al., Nature* 317: 230-234 (1985); Thomas *et al., Cell* 51: 503-512 (1987); Thompson *et al., Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g., Thomas, supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g., knockouts*) are administered to a patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC

compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt
5 the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

10 The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

15 Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.*, U.S. Patents 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

20 When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the
25 introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such
30 conditions and/or disorders.

Computer Readable Means

A further aspect of the invention relates to a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 1 through 136 and SEQ ID NO: 137 through 240 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set

representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

Diagnostic Methods for Prostate Cancer

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by

comparing expression of a PSNA or a PSP in a human patient that has or may have prostate cancer, or who is at risk of developing prostate cancer, with the expression of a PSNA or a PSP in a normal human control. For purposes of the present invention, “expression of a PSNA” or “PSNA expression” means the quantity of PSG mRNA that
5 can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term “expression of a PSP” or “PSP expression” means the amount of PSP that can be measured by any method known in the art or the level of translation of a PSG PSNA that can be measured by any method known in the art.

10 The present invention provides methods for diagnosing prostate cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of PSNA or PSP in cells, tissues, organs or bodily fluids compared with levels of PSNA or PSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of a PSNA or
15 PSP in the patient versus the normal human control is associated with the presence of prostate cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing prostate cancer in a patient by analyzing changes in the structure of the mRNA of a PSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing, alterations
20 in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing prostate cancer in a patient by analyzing changes in a PSP compared to a PSP from a normal control. These changes include, *e.g.*, alterations in glycosylation and/or phosphorylation of the PSP or subcellular PSP localization.

25 In a preferred embodiment, the expression of a PSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from SEQ ID NO: 137 through 240, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the PSNA expression that is measured is the level of expression of a PSNA mRNA selected from SEQ ID NO: 1 through 136, or a hybridizing nucleic
30 acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acids. PSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by Northern blot, quantitative

or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. See, e.g., Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. PSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of a PSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, e.g., aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, PSNA expression may be compared to a known control, such as normal prostate nucleic acid, to detect a change in expression.

10 In another preferred embodiment, the expression of a PSP is measured by determining the level of a PSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 137 through 240, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of PSNA or PSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of prostate cancer. The expression level of a PSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the PSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. See, e.g., Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the PSP structure may be determined by any method known in the art, including, e.g., using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

25 In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to a PSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-PSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a

protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the PSP will bind to the anti-PSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-PSP antibody that is linked to a detectable reagent (a radioactive
5 substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the PSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of a PSP in the sample. For an RIA, the solid support is counted for radioactive
10 decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure PSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-PSP antibody is attached to a solid support and an allocated amount of a labeled PSP and a sample of interest are incubated
15 with the solid support. The amount of labeled PSP detected which is attached to the solid support can be correlated to the quantity of a PSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides
20 are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with
25 chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of a PSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant
30 cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other

mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more PSNAs of interest. In this approach, all or a portion of one or more PSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of PSNA or PSP includes, without limitation, prostate tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, prostate cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary prostate cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and colon. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediastinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. *See Scott, supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in PSNAs or PSPs in cells in sputum samples may be particularly

useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of a PSNA or PSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other PSNA or PSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular PSNA or PSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more PSNAs and/or PSPs in a sample from a patient suspected of having prostate cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of a PSNA and/or PSP and then ascertaining whether the patient has prostate cancer from the expression level of the PSNA or PSP. In general, if high expression relative to a control of a PSNA or PSP is indicative of prostate cancer, a diagnostic assay is considered positive if the level of expression of the PSNA or PSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a PSNA or PSP is indicative of prostate cancer, a diagnostic assay is considered positive if the level of expression of the PSNA or PSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether prostate cancer has metastasized in a patient. One may identify whether the prostate cancer has

metastasized by measuring the expression levels and/or structural alterations of one or more PSNAs and/or PSPs in a variety of tissues. The presence of a PSNA or PSP in a certain tissue at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of a PSNA or PSP is associated with prostate cancer. Similarly, the presence of a PSNA or PSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of a PSNA or PSP is associated with prostate cancer. Further, the presence of a structurally altered PSNA or PSP that is associated with prostate cancer is also indicative of metastasis.

In general, if high expression relative to a control of a PSNA or PSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the PSNA or PSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a PSNA or PSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the PSNA or PSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

The PSNA or PSP of this invention may be used as element in an array or a multi-analyte test to recognize expression patterns associated with prostate cancers or other prostate related disorders. In addition, the sequences of either the nucleic acids or proteins may be used as elements in a computer program for pattern recognition of prostate disorders.

25

Staging

The invention also provides a method of staging prostate cancer in a human patient. The method comprises identifying a human patient having prostate cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more PSNAs or PSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the expression level of one or more PSNAs or PSPs is determined for each stage to

30

obtain a standard expression level for each PSNA and PSP. Then, the PSNA or PSP expression levels are determined in a biological sample from a patient whose stage of cancer is not known. The PSNA or PSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the

5 PSNAs and PSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of a PSNA or PSP to determine the stage of a prostate cancer.

Monitoring

Further provided is a method of monitoring prostate cancer in a human patient.

10 One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the prostate cancer. The method

15 comprises identifying a human patient that one wants to monitor for prostate cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more PSNAs or PSPs, and comparing the PSNA or PSP levels over time to those PSNA or PSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in a PSNA or PSP

20 that are associated with prostate cancer.

If increased expression of a PSNA or PSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of a PSNA or PSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One

25 having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of a PSNA or PSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an decrease in the expression level of a PSNA or PSP

30 indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of PSNAs or PSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples.

Monitoring a patient for onset of prostate cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of a PSNA and/or PSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more PSNAs and/or PSPs are detected. The presence of higher (or lower) PSNA or PSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly prostate cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more PSNAs and/or PSPs of the invention can also be monitored by analyzing levels of expression of the PSNAs and/or PSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

Detection of Genetic Lesions or Mutations

The methods of the present invention can also be used to detect genetic lesions or mutations in a PSG, thereby determining if a human with the genetic lesion is susceptible to developing prostate cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing prostate cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the PSGs of this invention, a chromosomal rearrangement of PSG, an aberrant modification of PSG (such as of the methylation pattern of the genomic DNA), or allelic loss of a PSG. Methods to detect such lesions in the PSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

Methods of Detecting Noncancerous Prostate Diseases

The invention also provides a method for determining the expression levels and/or structural alterations of one or more PSNAs and/or PSPs in a sample from a patient suspected of having or known to have a noncancerous prostate disease. In general, the method comprises the steps of obtaining a sample from the patient,

determining the expression level or structural alterations of a PSNA and/or PSP, comparing the expression level or structural alteration of the PSNA or PSP to a normal prostate control, and then ascertaining whether the patient has a noncancerous prostate disease. In general, if high expression relative to a control of a PSNA or PSP is

5 indicative of a particular noncancerous prostate disease, a diagnostic assay is considered positive if the level of expression of the PSNA or PSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a PSNA or PSP is

10 indicative of a noncancerous prostate disease, a diagnostic assay is considered positive if the level of expression of the PSNA or PSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same

15 patient.

One having ordinary skill in the art may determine whether a PSNA and/or PSP is associated with a particular noncancerous prostate disease by obtaining prostate tissue from a patient having a noncancerous prostate disease of interest and determining which PSNAs and/or PSPs are expressed in the tissue at either a higher or a lower level than in

20 normal prostate tissue. In another embodiment, one may determine whether a PSNA or PSP exhibits structural alterations in a particular noncancerous prostate disease state by obtaining prostate tissue from a patient having a noncancerous prostate disease of interest and determining the structural alterations in one or more PSNAs and/or PSPs relative to normal prostate tissue.

25 Methods for Identifying Prostate Tissue

In another aspect, the invention provides methods for identifying prostate tissue. These methods are particularly useful in, *e.g.*, forensic science, prostate cell

30 differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is prostate tissue or has prostate tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising prostate tissue or

having prostate tissue-like characteristics, determining whether the sample expresses one or more PSNAs and/or PSPs, and, if the sample expresses one or more PSNAs and/or PSPs, concluding that the sample comprises prostate tissue. In a preferred embodiment, the PSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 137 through 240, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the PSNA has a nucleotide sequence selected from SEQ ID NO: 1 through 137, or a hybridizing nucleic acid, an allelic variant or a part thereof.

Determining whether a sample expresses a PSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether a PSP is expressed. Determining whether a sample expresses a PSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the PSP has an amino acid sequence selected from SEQ ID NO: 137 through 240, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two PSNAs and/or PSPs is determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five PSNAs and/or PSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is prostate tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into prostate tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new prostate tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

Methods for Producing and Modifying Prostate Tissue

In another aspect, the invention provides methods for producing engineered prostate tissue or cells. In one embodiment, the method comprises the steps of providing
5 cells, introducing a PSNA or a PSG into the cells, and growing the cells under conditions in which they exhibit one or more properties of prostate tissue cells. In a preferred embodiment, the cells are pluripotent. As is well-known in the art, normal prostate tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered prostate tissue or cells comprises one of these cell types. In another
10 embodiment, the engineered prostate tissue or cells comprises more than one prostate cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the prostate cell tissue. Methods for manipulating culture conditions are well-known in the art.

Nucleic acid molecules encoding one or more PSPs are introduced into cells,
15 preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode PSPs having amino acid sequences selected from SEQ ID NO: 137 through 240, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1 through 136, or hybridizing nucleic acids, allelic variants or parts
20 thereof. In another highly preferred embodiment, a PSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

Artificial prostate tissue may be used to treat patients who have lost some or all of their prostate function.

25 Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a
30 preferred embodiment, the pharmaceutical composition comprises a PSNA or part thereof. In a more preferred embodiment, the PSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 136, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity

thereto. In another preferred embodiment, the pharmaceutical composition comprises a PSP or fragment thereof. In a more preferred embodiment, the PSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 137 through 240, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-PSP antibody, preferably an antibody that specifically binds to a PSP having an amino acid that is selected from the group consisting of SEQ ID NO: 137 through 240, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art, and is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000), the disclosures of which are incorporated herein by reference in their entireties, and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium

carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

5 Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

 Tablet binders that can be used include acacia, methylcellulose, sodium
10 carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

 Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

 Fillers, agents that facilitate disintegration and/or solubilization, tablet binders
15 and lubricants, including the aforementioned, can be used singly or in combination.

 Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic
20 solvents or solvent mixtures.

 Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and,
25 optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

 Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

30 Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin,

carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

- 5 The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 10 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

- Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a 15 pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as 20 sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

- 25 Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

- 30 Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as

polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that
5 are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically.

For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and
10 can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin
15 penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases
20 as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various
25 powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for
30 reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding
5 free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of
10 those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example PSP polypeptide, fusion protein, or fragments thereof, antibodies specific for PSP, agonists, antagonists or inhibitors of PSP, which ameliorates the signs or symptoms of the disease or prevents progression thereof, as would be understood in the medical
15 arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of
20 administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical
25 compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of
30 active agent varies within this range depending upon pharmacokinetic factors well-known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

-111-

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug

5 combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic
10 agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the
15 literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine,
20 can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

Therapeutic Methods

25 The present invention further provides methods of treating subjects having defects in a gene of the invention, *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of prostate function. As used herein, "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any
30 improvement of a disease, including minor improvements. These methods are discussed below.

Gene Therapy and Vaccines

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication
5 incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV), for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of "naked" nucleic acid vaccination, as further described in U.S. Patents 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913;
10 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; and 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g., Doronin et al., J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a
15 therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of a PSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of a PSP are administered, for example, to complement a deficiency in the native PSP, or as DNA
20 vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g., Cid-Arregui, supra.* In a preferred embodiment, the nucleic acid molecule encodes a PSP having the amino acid sequence of SEQ ID NO: 137 through 240, or a fragment, fusion protein, allelic variant or homolog thereof.

25 In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express a PSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in PSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells
30 encode a PSP having the amino acid sequence of SEQ ID NO: 137 through 240, or a fragment, fusion protein, allelic variant or homolog thereof.

Antisense Administration

Antisense nucleic acid compositions, or vectors that drive expression of a PSG antisense nucleic acid, are administered to downregulate transcription and/or translation of a PSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of a PSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to PSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3):97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the PSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding a PSP, preferably a PSP comprising an amino acid sequence of SEQ ID NO: 137 through 240, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 136, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Polypeptide Administration

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a PSP, a

fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant PSP defect.

Protein compositions are administered, for example, to complement a deficiency in native PSP. In other embodiments, protein compositions are administered as a vaccine
5 to elicit a humoral and/or cellular immune response to PSP. The immune response can be used to modulate activity of PSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate PSP.

10 In a preferred embodiment, the polypeptide is a PSP comprising an amino acid sequence of SEQ ID NO: 137 through 240, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1
15 through 136, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is
20 administered. As is well-known, antibody compositions are administered, for example, to antagonize activity of PSP, or to target therapeutic agents to sites of PSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to a PSP comprising an amino acid sequence of SEQ ID NO: 137 through 240, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred
25 embodiment, the antibody specifically binds to a PSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 136, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to a PSP or have a modulatory effect on the expression or activity of a PSP.
30 Modulators which decrease the expression or activity of PSP (antagonists) are believed to be useful in treating prostate cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small

molecules predicted via computer imaging to specifically bind to regions of a PSP can also be designed, synthesized and tested for use in the imaging and treatment of prostate cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the PSPs identified herein. Molecules
5 identified in the library as being capable of binding to a PSP are key candidates for further evaluation for use in the treatment of prostate cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of a PSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of PSP is
10 administered. Antagonists of PSP can be produced using methods generally known in the art. In particular, purified PSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of a PSP.

In other embodiments a pharmaceutical composition comprising an agonist of a
15 PSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a PSP comprising an amino acid sequence of SEQ ID NO: 137 through 240, or a fusion protein, allelic variant, homolog, analog or
20 derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a PSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 136, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting Prostate Tissue

25 The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the prostate or to specific cells in the prostate. In a preferred embodiment, an anti-PSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if prostate tissue needs to be
30 selectively destroyed. This would be useful for targeting and killing prostate cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting prostate cell function.

In another embodiment, an anti-PSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring prostate function, identifying prostate cancer tumors, and identifying noncancerous prostate diseases.

5

EXAMPLES

Example 1: Gene Expression analysis

PSGs were identified by a systematic analysis of gene expression data in the LIFESEQ® Gold database available from Incyte Genomics Inc (Palo Alto, CA) using the data mining software package CLASP™ (Candidate Lead Automatic Search Program). CLASP™ is a set of algorithms that interrogate Incyte's database to identify genes that are both specific to particular tissue types as well as differentially expressed in tissues from patients with cancer. LifeSeq® Gold contains information about which genes are expressed in various tissues in the body and about the dynamics of expression in both normal and diseased states. CLASP™ first sorts the LifeSeq® Gold database into defined tissue types, such as breast, ovary and prostate. CLASP™ categorizes each tissue sample by disease state. Disease states include "healthy," "cancer," "associated with cancer," "other disease" and "other." Categorizing the disease states improves our ability to identify tissue and cancer-specific molecular targets. CLASP™ then performs a simultaneous parallel search for genes that are expressed both (1) selectively in the defined tissue type compared to other tissue types and (2) differentially in the "cancer" disease state compared to the other disease states affecting the same, or different, tissues. This sorting is accomplished by using mathematical and statistical filters that specify the minimum change in expression levels and the minimum frequency that the differential expression pattern must be observed across the tissue samples for the gene to be considered statistically significant. The CLASP™ algorithm quantifies the relative abundance of a particular gene in each tissue type and in each disease state.

To find the PSGs of this invention, the following specific CLASP™ profiles were utilized: tissue-specific expression (CLASP 1), detectable expression only in cancer tissue (CLASP 2), highest differential expression for a given cancer (CLASP 4); differential expression in cancer tissue (CLASP 5), and. cDNA libraries were divided

30

into 60 unique tissue types (early versions of LifeSeq® had 48 tissue types). Genes or ESTs were grouped into “gene bins,” where each bin is a cluster of sequences grouped together where they share a common contig. The expression level for each gene bin was calculated for each tissue type. Differential expression significance was calculated with rigorous statistical significant testing taking into account variations in sample size and relative gene abundance in different libraries and within each library (for the equations used to determine statistically significant expression see Audic and Claverie “The significance of digital gene expression profiles,” *Genome Res* 7(10): 986-995 (1997), including Equation 1 on page 987 and Equation 2 on page 988, the contents of which are incorporated by reference). Differentially expressed tissue-specific genes were selected based on the percentage abundance level in the targeted tissue versus all the other tissues (tissue-specificity). The expression levels for each gene in libraries of normal tissues or non-tumor tissues from cancer patients were compared with the expression levels in tissue libraries associated with tumor or disease (cancer-specificity). The results were analyzed for statistical significance.

The selection of the target genes meeting the rigorous CLASP™ profile criteria were as follows:

- (a) CLASP 1: tissue-specific expression: To qualify as a CLASP 1 candidate, a gene must exhibit statistically significant expression in the tissue of interest compared to all other tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 1 candidate.
- (b) CLASP 2: detectable expression only in cancer tissue: To qualify as a CLASP 2 candidate, a gene must exhibit detectable expression in tumor tissues and undetectable expression in libraries from normal individuals and libraries from normal tissue obtained from diseased patients. In addition, such a gene must also exhibit further specificity for the tumor tissues of interest.
- (c) CLASP 5: differential expression in cancer tissue: To qualify as a CLASP 5 candidate, a gene must be differentially expressed in tumor libraries in the tissue of interest compared to normal libraries for all tissues. Only if the gene exhibits such differential expression with a 90% of confidence level is it selected as a CLASP 5 candidate.

The CLASP™ scores for SEQ ID NO: 1-136 are listed below:

The CLASP™ scores for SEQ ID NO: 1-136 are listed below:

	SEQ ID NO: 1	DEX0259_1	CLASP2
5	SEQ ID NO: 2	DEX0259_2	CLASP2
	SEQ ID NO: 3	DEX0259_3	CLASP2 CLASP1
	SEQ ID NO: 4	DEX0259_4	CLASP2 CLASP1
	SEQ ID NO: 5	DEX0259_5	CLASP2
	SEQ ID NO: 6	DEX0259_6	CLASP2
10	SEQ ID NO: 7	DEX0259_7	CLASP2
	SEQ ID NO: 8	DEX0259_8	CLASP2
	SEQ ID NO: 9	DEX0259_9	CLASP2
	SEQ ID NO: 10	DEX0259_10	CLASP2 CLASP1
	SEQ ID NO: 11	DEX0259_11	CLASP2 CLASP1
15	SEQ ID NO: 12	DEX0259_12	CLASP2
	SEQ ID NO: 13	DEX0259_13	CLASP2
	SEQ ID NO: 14	DEX0259_14	CLASP5 CLASP1
	SEQ ID NO: 15	DEX0259_15	CLASP5
	SEQ ID NO: 16	DEX0259_16	CLASP5 CLASP1
20	SEQ ID NO: 17	DEX0259_17	CLASP5 CLASP1
	SEQ ID NO: 18	DEX0259_18	CLASP2
	SEQ ID NO: 20	DEX0259_20	CLASP2
	SEQ ID NO: 21	DEX0259_21	CLASP2
	SEQ ID NO: 22	DEX0259_22	CLASP2
25	SEQ ID NO: 23	DEX0259_23	CLASP2
	SEQ ID NO: 24	DEX0259_24	CLASP2
	SEQ ID NO: 25	DEX0259_25	CLASP2
	SEQ ID NO: 26	DEX0259_26	CLASP5 CLASP1
	SEQ ID NO: 27	DEX0259_27	CLASP5 CLASP1
30	SEQ ID NO: 28	DEX0259_28	CLASP2
	SEQ ID NO: 29	DEX0259_29	CLASP1
	SEQ ID NO: 30	DEX0259_30	CLASP5 CLASP1
	SEQ ID NO: 31	DEX0259_31	CLASP5 CLASP1
	SEQ ID NO: 32	DEX0259_32	CLASP5 CLASP1
35	SEQ ID NO: 33	DEX0259_33	CLASP2
	SEQ ID NO: 34	DEX0259_34	CLASP2
	SEQ ID NO: 35	DEX0259_35	CLASP2
	SEQ ID NO: 36	DEX0259_36	CLASP2
	SEQ ID NO: 37	DEX0259_37	CLASP2
40	SEQ ID NO: 38	DEX0259_38	CLASP2
	SEQ ID NO: 39	DEX0259_39	CLASP2
	SEQ ID NO: 40	DEX0259_40	CLASP2
	SEQ ID NO: 41	DEX0259_41	CLASP2
	SEQ ID NO: 42	DEX0259_42	CLASP5
45	SEQ ID NO: 43	DEX0259_43	CLASP1
	SEQ ID NO: 44	DEX0259_44	CLASP1

	SEQ ID NO: 45	DEX0259_45 CLASP5 CLASP1
	SEQ ID NO: 46	DEX0259_46 CLASP2
	SEQ ID NO: 47	DEX0259_47 CLASP2 CLASP1
	SEQ ID NO: 48	DEX0259_48 CLASP2
5	SEQ ID NO: 49	DEX0259_49 CLASP2
	SEQ ID NO: 50	DEX0259_50 CLASP2
	SEQ ID NO: 51	DEX0259_51 CLASP5 CLASP1
	SEQ ID NO: 53	DEX0259_53 CLASP2
	SEQ ID NO: 54	DEX0259_54 CLASP2
10	SEQ ID NO: 55	DEX0259_55 CLASP2
	SEQ ID NO: 56	DEX0259_56 CLASP2
	SEQ ID NO: 57	DEX0259_57 CLASP2
	SEQ ID NO: 58	DEX0259_58 CLASP2
	SEQ ID NO: 59	DEX0259_59 CLASP2
15	SEQ ID NO: 60	DEX0259_60 CLASP2
	SEQ ID NO: 61	DEX0259_61 CLASP2
	SEQ ID NO: 62	DEX0259_62 CLASP2
	SEQ ID NO: 63	DEX0259_63 CLASP1
	SEQ ID NO: 64	DEX0259_64 CLASP1
20	SEQ ID NO: 65	DEX0259_65 CLASP2 CLASP1
	SEQ ID NO: 66	DEX0259_66 CLASP2 CLASP1
	SEQ ID NO: 67	DEX0259_67 CLASP2
	SEQ ID NO: 68	DEX0259_68 CLASP2
	SEQ ID NO: 71	DEX0259_71 CLASP2
25	SEQ ID NO: 72	DEX0259_72 CLASP2
	SEQ ID NO: 73	DEX0259_73 CLASP2
	SEQ ID NO: 74	DEX0259_74 CLASP2
	SEQ ID NO: 75	DEX0259_75 CLASP2
	SEQ ID NO: 76	DEX0259_76 CLASP2
30	SEQ ID NO: 77	DEX0259_77 CLASP2
	SEQ ID NO: 78	DEX0259_78 CLASP2
	SEQ ID NO: 79	DEX0259_79 CLASP2
	SEQ ID NO: 80	DEX0259_80 CLASP2
	SEQ ID NO: 81	DEX0259_81 CLASP2
35	SEQ ID NO: 82	DEX0259_82 CLASP2 CLASP1
	SEQ ID NO: 83	DEX0259_83 CLASP2
	SEQ ID NO: 84	DEX0259_84 CLASP2
	SEQ ID NO: 85	DEX0259_85 CLASP2
	SEQ ID NO: 86	DEX0259_86 CLASP2
40	SEQ ID NO: 87	DEX0259_87 CLASP2
	SEQ ID NO: 88	DEX0259_88 CLASP2
	SEQ ID NO: 89	DEX0259_89 CLASP2 CLASP1
	SEQ ID NO: 90	DEX0259_90 CLASP2
	SEQ ID NO: 91	DEX0259_91 CLASP2
45	SEQ ID NO: 92	DEX0259_92 CLASP2
	SEQ ID NO: 93	DEX0259_93 CLASP2
	SEQ ID NO: 94	DEX0259_94 CLASP2
	SEQ ID NO: 95	DEX0259_95 CLASP2

-120-

	SEQ ID NO: 96	DEX0259_96	CLASP2
	SEQ ID NO: 97	DEX0259_97	CLASP2
	SEQ ID NO: 98	DEX0259_98	CLASP2
	SEQ ID NO: 99	DEX0259_99	CLASP2
5	SEQ ID NO: 100	DEX0259_100	CLASP2
	SEQ ID NO: 101	DEX0259_101	CLASP2
	SEQ ID NO: 102	DEX0259_102	CLASP2
	SEQ ID NO: 103	DEX0259_103	CLASP2
	SEQ ID NO: 104	DEX0259_104	CLASP2
10	SEQ ID NO: 105	DEX0259_105	CLASP2
	SEQ ID NO: 106	DEX0259_106	CLASP2
	SEQ ID NO: 107	DEX0259_107	CLASP2
	SEQ ID NO: 108	DEX0259_108	CLASP5
	SEQ ID NO: 109	DEX0259_109	CLASP2
15	SEQ ID NO: 110	DEX0259_110	CLASP2
	SEQ ID NO: 111	DEX0259_111	CLASP2
	SEQ ID NO: 112	DEX0259_112	CLASP2
	SEQ ID NO: 113	DEX0259_113	CLASP2
	SEQ ID NO: 114	DEX0259_114	CLASP2
20	SEQ ID NO: 115	DEX0259_115	CLASP2 CLASP1
	SEQ ID NO: 116	DEX0259_116	CLASP2 CLASP1
	SEQ ID NO: 117	DEX0259_117	CLASP2 CLASP1
	SEQ ID NO: 118	DEX0259_118	CLASP2 CLASP1
	SEQ ID NO: 119	DEX0259_119	CLASP2
25	SEQ ID NO: 120	DEX0259_120	CLASP2 CLASP1
	SEQ ID NO: 121	DEX0259_121	CLASP2
	SEQ ID NO: 122	DEX0259_122	CLASP2 CLASP1
	SEQ ID NO: 123	DEX0259_123	CLASP2 CLASP1
	SEQ ID NO: 124	DEX0259_124	CLASP2
30	SEQ ID NO: 125	DEX0259_125	CLASP2
	SEQ ID NO: 126	DEX0259_126	CLASP2
	SEQ ID NO: 127	DEX0259_127	CLASP2
	SEQ ID NO: 128	DEX0259_128	CLASP2
	SEQ ID NO: 129	DEX0259_129	CLASP2
35	SEQ ID NO: 130	DEX0259_130	CLASP2
	SEQ ID NO: 131	DEX0259_131	CLASP2
	SEQ ID NO: 132	DEX0259_132	CLASP2
	SEQ ID NO: 133	DEX0259_133	CLASP2
	SEQ ID NO: 134	DEX0259_134	CLASP2
40	SEQ ID NO: 135	DEX0259_135	CLASP2
	SEQ ID NO: 136	DEX0259_136	CLASP2

DEX0259 CLASP expression Level

45	SEQ ID NO: 1PRO .0019			
	SEQ ID NO: 2PRO .0038			
	SEQ ID NO: 3PRO .0071	FTS .0001	BLO .0003	INL .0004
	SEQ ID NO: 4PRO .0071	FTS .0001	BLO .0003	INL .0004

-121-

SEQ ID NO: 5PRO .0038
 SEQ ID NO: 6PRO .0038
 SEQ ID NO: 7PRO .0038
 SEQ ID NO: 8PRO .002
 5 SEQ ID NO: 10 PRO .0051
 SEQ ID NO: 11 PRO .0051
 SEQ ID NO: 12 PRO .002 UTR .0056 URE .0117
 SEQ ID NO: 13 PRO .002 UTR .0056 URE .0117
 SEQ ID NO: 14 PRO .0017 FTS .0001
 10 SEQ ID NO: 15 PRO .1096 ADR .0376 BRN .0488 PNS .0491 KID .0781
 SEQ ID NO: 16 PRO .0017 FTS .0001 INL .0004 NRV .0009
 SEQ ID NO: 17 PRO .0017 FTS .0001 INL .0004 NRV .0009
 SEQ ID NO: 18 PRO .0039
 SEQ ID NO: 20 PRO .0026
 15 SEQ ID NO: 21 PRO .0026
 SEQ ID NO: 22 PRO .002 INL .0025
 SEQ ID NO: 23 PRO .002 INL .0025
 SEQ ID NO: 24 PRO .0044
 SEQ ID NO: 25 PRO .0044
 20 SEQ ID NO: 26 PRO .004 FTS .0001 BLO .0003 KID .0013
 SEQ ID NO: 27 PRO .004 FTS .0001 BLO .0003 KID .0013
 SEQ ID NO: 28 PRO .0032
 SEQ ID NO: 29 PRO .0011
 SEQ ID NO: 30 PRO .0017 MAM .0004
 25 SEQ ID NO: 31 PRO .0017 FTS .0001
 SEQ ID NO: 32 PRO .0684 BNC .0031 NRV .0035 KID .0039 FTS .0039
 SEQ ID NO: 33 PRO .002
 SEQ ID NO: 34 PRO .002
 SEQ ID NO: 35 PRO .0032
 30 SEQ ID NO: 36 PRO .0032
 SEQ ID NO: 37 PRO .002
 SEQ ID NO: 38 PRO .002
 SEQ ID NO: 39 PRO .0017
 SEQ ID NO: 40 PRO .0017
 35 SEQ ID NO: 41 PRO .0017
 SEQ ID NO: 42 PRO .0011
 SEQ ID NO: 43 PRO .0011 FTS .0003 STO .0021
 SEQ ID NO: 44 PRO .0011 FTS .0003 STO .0021
 SEQ ID NO: 45 PRO .0017
 40 SEQ ID NO: 46 PRO .0021
 SEQ ID NO: 47 PRO .0042
 SEQ ID NO: 48 PRO .0021
 SEQ ID NO: 49 PRO .0021
 SEQ ID NO: 50 PRO .0021 INL .0025
 45 SEQ ID NO: 51 PRO .0017 MAM .0008 NRV .0009
 SEQ ID NO: 53 PRO .0021
 SEQ ID NO: 54 PRO .0021
 SEQ ID NO: 55 PRO .0021

-122-

	SEQ ID NO: 56	PRO .0021			
	SEQ ID NO: 57	PRO .0021			
	SEQ ID NO: 58	PRO .0021			
	SEQ ID NO: 59	PRO .0021			
5	SEQ ID NO: 60	PRO .0021			
	SEQ ID NO: 61	PRO .0021			
	SEQ ID NO: 62	PRO .0021			
	SEQ ID NO: 63	PRO .0017 LNG .0004	UTR .0004	INL .0004	BLD .0016
	SEQ ID NO: 64	PRO .0017 LNG .0004	UTR .0004	INL .0004	BLD .0016
10	SEQ ID NO: 65	PRO .0039			
	SEQ ID NO: 66	PRO .0039			
	SEQ ID NO: 67	PRO .0013	CRD .0138		
	SEQ ID NO: 68	PRO .0013	CRD .0138		
	SEQ ID NO: 71	PRO .0013			
15	SEQ ID NO: 72	PRO .0013			
	SEQ ID NO: 73	PRO .0013			
	SEQ ID NO: 74	PRO .0013			
	SEQ ID NO: 75	PRO .0013			
	SEQ ID NO: 76	PRO .0013			
20	SEQ ID NO: 77	PRO .0013			
	SEQ ID NO: 78	PRO .0013			
	SEQ ID NO: 79	PRO .0013			
	SEQ ID NO: 80	PRO .0013			
	SEQ ID NO: 81	PRO .0013			
25	SEQ ID NO: 82	PRO .0013			
	SEQ ID NO: 83	PRO .0013			
	SEQ ID NO: 84	PRO .0013			
	SEQ ID NO: 85	PRO .0013			
	SEQ ID NO: 86	PRO .0013			
30	SEQ ID NO: 87	PRO .0013			
	SEQ ID NO: 88	PRO .0013			
	SEQ ID NO: 89	PRO .002			
	SEQ ID NO: 90	PRO .0013			
	SEQ ID NO: 91	PRO .0013			
35	SEQ ID NO: 92	PRO .0013			
	SEQ ID NO: 93	PRO .0013			
	SEQ ID NO: 94	PRO .0013			
	SEQ ID NO: 95	PRO .0013			
	SEQ ID NO: 96	PRO .002			
40	SEQ ID NO: 97	PRO .002			
	SEQ ID NO: 98	PRO .002			
	SEQ ID NO: 99	PRO .002			
	SEQ ID NO: 100	PRO .002			
	SEQ ID NO: 101	PRO .002			
45	SEQ ID NO: 102	PRO .002			
	SEQ ID NO: 103	PRO .002			
	SEQ ID NO: 104	PRO .002			
	SEQ ID NO: 105	PRO .002			

-123-

	SEQ ID NO: 106	PRO .002	
	SEQ ID NO: 107	PRO .002	
	SEQ ID NO: 108	PRO .0006	
	SEQ ID NO: 109	PRO .0042	
5	SEQ ID NO: 110	PRO .0042	
	SEQ ID NO: 111	PRO .002	LNG .0015
	SEQ ID NO: 112	PRO .002	LNG .0015
	SEQ ID NO: 113	PRO .002	
	SEQ ID NO: 114	PRO .002	
10	SEQ ID NO: 115	PRO .004	
	SEQ ID NO: 116	PRO .004	
	SEQ ID NO: 117	PRO .003	
	SEQ ID NO: 118	PRO .003	
	SEQ ID NO: 119	PRO .002	
15	SEQ ID NO: 120	PRO .003	
	SEQ ID NO: 121	PRO .002	
	SEQ ID NO: 122	PRO .002	INL .0004
	SEQ ID NO: 123	PRO .002	INL .0004
	SEQ ID NO: 124	PRO .002	
20	SEQ ID NO: 125	PRO .002	
	SEQ ID NO: 126	PRO .002	
	SEQ ID NO: 127	PRO .002	
	SEQ ID NO: 128	PRO .002	
	SEQ ID NO: 129	PRO .002	
25	SEQ ID NO: 130	PRO .002	
	SEQ ID NO: 131	PRO .002	
	SEQ ID NO: 132	PRO .002	
	SEQ ID NO: 133	PRO .002	
	SEQ ID NO: 134	PRO .0039	
30	SEQ ID NO: 135	PRO .0039	INS .0128
	SEQ ID NO: 136	PRO .0039	INS .0128

Abbreviation for tissues:

BLO Blood; BRN Brain; CON Connective Tissue; CRD Heart; FTS Fetus; INL Intestine, Large; INS Intestine, Small; KID Kidney; LIV Liver; LNG Lung; MAM Breast; MSL
 35 Muscles; NRV Nervous Tissue; OVR Ovary; PRO Prostate; STO Stomach; THR
 Thyroid Gland; TNS Tonsil / Adenoids; UTR Uterus

Example 2: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation
 40 detection system utilizing the 5'-3' nuclease activity of Taq DNA polymerase. The
 method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5'
 reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity
 of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected
 by the laser detector of the Model 7700 Sequence Detection System (PE Applied

Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this
5 endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

10 The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman probes specific to each target gene. The
15 results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the PSNA versus normal tissues and other cancer tissues can then be
20 determined. All the values are compared to normal thymus (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the PSNA in pairs of matching samples and 1 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values
25 are compared to normal thymus (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

In the analysis of matching samples, the PSNAs that show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results
30 obtained with normal pooled samples.

Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an

-125-

indication of specificity for the cancer stage (*e.g.* higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1 through 136 being a diagnostic

5 marker for cancer.

Sequences	Sequence ID NO	Gene ID	QPCR prostate code
DEX0098_17	DEX0259_26(SEQ ID NO:26)	14049	Pro166
	DEX0259_27(SEQ ID NO:27)		
DEX0098_21	DEX0259_31(SEQ ID NO:31)	146117	Pro162
10	DEX0259_32(SEQ ID NO:32)		

Sequence ID NO: DEX0259_26(SEQ ID NO:26) DEX0259_27(SEQ ID NO:27)
Pro166

15 QPCR data was inconclusive.

Primers Used for QPCR Expression Analysis
In DEX0259_26(SEQ ID NO:26)

Primer Probe Oligo	Start From	End To	queryLength	sbjctDescript
Pro166For	211	235	25	DEX0098_17
Pro166Rev	347	323	25	DEX0098_17
Pro166Probe	308	274	35	DEX0098_17

20

In DEX0259_27(SEQ ID NO:27)

Primer Probe Oligo	Start From	End To	Query Length	sbjctDescript
Pro166For	3256	3232	25	flexsednt DEX0098_17
Pro166Rev	3120	3144	25	flexsednt DEX0098_17
Pro166Probe	3159	3193	35	flexsednt DEX0098_17

Experiments results from SQ PCR analysis are included below.

25 SQ code for Pro166: sqpro093

The relative levels of expression of Sqpro093 in 12 normal samples from 12 different tissues are listed below. These RNA samples are individual samples or are commercially available pools, originated by pooling samples of a particular tissue from different

30 individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of

-126-

0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Tissue	Normal
Breast	10
Colon	10
Endometrium	1
Kidney	10
Liver	0
Lung	1
Ovary	10
Prostate	1000
Small Intestine	100
Stomach	1
Testis	1
Uterus	1

- 5 Relative levels of expression in the table below shows that highest expression level of Sqpro093 is detected in prostate.

The relative levels of expression of Sqpro093 in 12 cancer samples from 12 different tissues are shown below. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate.

- 10 Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Tissue	Cancer
Bladder	1
Breast	1
Colon	1
Kidney	0
Liver	0
Lung	10
Ovary	10
Pancreas	10
Prostate	100
Stomach	10
Testis	100
Uterus	10

- 15 Relative levels of expression in Table 2 show that high expression level of Sqpro093 is detected in prostate and testis carcinomas.

-127-

The relative levels of expression of Sqpro093 in 6 prostate cancer matching samples are shown below. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

- 5 Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Sample ID	Tissue	Cancer	NAT
845B/846B	Prostate	10	100
916B/917B	Prostate	100	100
1105B/1106B	Prostate	100	10
902B/903B	Prostate	1000	100
1222B/1223B	Prostate	10	10
1291B/1292B	Prostate	100	10

10

Relative levels of expression in Table 3 shows that Sqpro093 is expressed in higher level in two of all six prostate cancer samples compared with their normal adjacent matching pair.

- 15 Sequence ID NO: DEX0259_31(SEQ ID NO:31) & DEX0259_32(SEQ ID NO:32)
Pro162

QPCR data was inconclusive.

In DEX0259_31(SEQ ID NO:31)

Primer Probe Oligo	Start From	End To	Query Length	sbjctDescript
Pro162For	344	364	21	DEX0098_21
Pro162Rev	481	459	23	DEX0098_21
Pro162Probe	435	403	33	DEX0098_21

20

In DEX0259_32(SEQ ID NO:32)

Primer Probe Oligo	Start From	End To	Query Length	sbjctDescript
Pro162For	344	364	21	flexsednt DEX0098_21
Pro162Rev	481	459	23	flexsednt DEX0098_21
Pro162Probe	435	403	33	flexsednt

				DEX0098 21
--	--	--	--	------------

Experimental results from SQ PCR analysis are included below.

SQ code for Pro161: sqpro076

- 5 The relative levels of expression of Sqpro076 in 12 normal samples from 12 different tissues are shown below. These RNA samples are individual samples or are commercially available pools, originated by pooling samples of a particular tissue from different individuals. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative
- 10 expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Tissue	Normal
Breast	0
Colon	0
Endometrium	0
Kidney	0
Liver	0
Lung	0
Ovary	0
Prostate	0
Small Intes\tine	0
Stomach	0
Testis	0
Uterus	0

Expression of sqpro076 is not detected in all 12 normal tissues.

- 15 The relative levels of expression of Sqpro076 in 12 cancer samples from 12 different tissues are shown below. Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression
- 20 value.

Tissue	Cancer
Bladder	0
Breast	0
Colon	0
Kidney	0
Liver	0

Lung	0
Ovary	0
Pancreas	0
Prostate	0
Stomach	0
Testis	0
Uterus	0

Relative levels of expression show that expression of Sqpro076 is not detected in all 12 carcinomas.

The relative levels of expression of Sqpro076 in 6 prostate cancer matching
 5 samples are shown below. A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

Using Polymerase Chain Reaction (PCR) technology expression levels were analyzed from four 10x serial cDNA dilutions in duplicate. Relative expression levels of
 10 0, 1, 10, 100 and 1000 are used to evaluate gene expression. A positive reaction in the most dilute sample indicates the highest relative expression value.

Sample ID	Tissue	Cancer	NAT
845B/846B	Prostate	1	1
916B/917B	Prostate	10	10
1105B/1106B	Prostate	10	10
902B/903B	Prostate	10	1
1222B/1223B	Prostate	100	1
1291B/1292B	Prostate	1	0

Relative levels of expression in Table 3 show that Sqpro076 is expressed in higher levels
 15 in three of the six prostate cancer samples compared with their normal adjacent matching pair.

Example 3: Protein Expression

The PSNA is amplified by polymerase chain reaction (PCR) and the amplified
 20 DNA fragment encoding the PSNA is subcloned in pET-21d for expression in *E. coli*. In addition to the PSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH₂-terminus of the coding sequence of PSNA, and six histidines, flanking the COOH-terminus of the coding sequence of PSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of PSP was achieved using cell paste generated from
5 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that had been separated from total cell lysate were incubated with a nickel chelating resin. The column was packed and washed with five column volumes of wash buffer. PSP was eluted stepwise with various concentration imidazole buffers.

10 **Example 4: Protein Fusions**

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4
15 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without
20 a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e. g.*, WO 96/34891.

25 **Example 5: Production of an Antibody from a Polypeptide**

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine
30 serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The

splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT
 5 medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be
 10 produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the
 15 hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies. Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and
 20 Wolf, CABIOS, 4(1), 181-186, 1988, the contents of which are incorporated by reference).

	DEX0259_141	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	17-30	1.06	14
25	DEX0259_143	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	85-99	1.13	15
30	DEX0259_146	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	49-83	1.10	35
	DEX0259_147	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	31-45	1.12	15
35	DEX0259_151	Antigenicity Index(Jameson-Wolf)	
	positions	AI	avg length
	16-27	1.01	12
	DEX0259_153	Antigenicity Index(Jameson-Wolf)	

-132-

		positions	AI	avg length
		804-820	1.23	17
		861-875	1.22	15
		9-26	1.14	18
5		200-212	1.07	13
		351-361	1.07	11
		636-646	1.07	11
		214-230	1.02	17
		599-620	1.01	22
10	DEX0259_158	Antigenicity Index(Jameson-Wolf)		
		positions	AI	avg length
		20-32	1.22	13
		172-185	1.11	14
		192-208	1.08	17
15		106-118	1.07	13
	DEX0259_162	Antigenicity Index(Jameson-Wolf)		
		positions	AI	avg length
		55-71	1.11	17
	DEX0259_167	Antigenicity Index(Jameson-Wolf)		
20		positions	AI	avg length
		20-36	1.22	17
	DEX0259_175	Antigenicity Index(Jameson-Wolf)		
		positions	AI	avg length
		14-26	1.27	13
25	DEX0259_193	Antigenicity Index(Jameson-Wolf)		
		positions	AI	avg length
		20-36	1.00	17
	DEX0259_198	Antigenicity Index(Jameson-Wolf)		
		positions	AI	avg length
30		118-129	1.16	12
	DEX0259_206	Antigenicity Index(Jameson-Wolf)		
		positions	AI	avg length
		14-27	1.06	14
	DEX0259_223	Antigenicity Index(Jameson-Wolf)		
35		positions	AI	avg length
		10-19	1.19	10
	DEX0259_224	Antigenicity Index(Jameson-Wolf)		
		positions	AI	avg length
		66-83	1.11	18
40		540-550	1.11	11
		204-217	1.09	14
		696-710	1.02	15
		418-443	1.01	26
	DEX0259_238	Antigenicity Index(Jameson-Wolf)		
45		positions	AI	avg length
		15-39	1.22	25

Examples of post-translational modifications (PTMs) of the PSPs of this invention are listed below. In addition, antibodies that specifically bind such post-translational modifications may be useful as a diagnostic or as therapeutic. Using the ProSite database (Bairoch et al., Nucleic Acids Res. 25(1):217-221 (1997), the contents of which are incorporated by reference), the following PTMs were predicted for the PSPs of the invention (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_prosite.html most recently accessed October 23, 2001). For full definitions of the PTMs see <http://www.expasy.org/cgi-bin/prosite-list.pl> most recently accessed October 23, 2001.

	DEX0259_137	Pkc_Phospho_Site 24-26;
10	DEX0259_141	Asn_Glycosylation 45-48; Ck2_Phospho_Site 12-15;25-28; Myristyl 61-66; Pkc_Phospho_Site 18-20;
	DEX0259_143	Asn_Glycosylation 17-20;93-96;128-131;145-148;164-167;196-199; Ck2_Phospho_Site 35-38;50-53;191-194; Myristyl 69-74;92-97; Pkc_Phospho_Site 57-59;87-89;97-99;123-125;137-139;182-184;198-200;
15	DEX0259_144	Amidation 59-62; Camp_Phospho_Site 29-32; Ck2_Phospho_Site 32-35; Myristyl 54-59;75-80;79-84; Pkc_Phospho_Site 32-34;
	DEX0259_145	Myristyl 15-20; Pkc_Phospho_Site 19-21;
	DEX0259_146	Ck2_Phospho_Site 47-50; Myristyl 6-11;9-14;43-48;91-96;
20		Pkc_Phospho_Site 18-20;
	DEX0259_147	Asn_Glycosylation 41-44;62-65; Pkc_Phospho_Site 38-40;44-46;
	DEX0259_148	Asn_Glycosylation 6-9; Ck2_Phospho_Site 8-11;
		Pkc_Phospho_Site 5-7;
	DEX0259_150	Myristyl 33-38;
25	DEX0259_151	Myristyl 19-24;33-38;
	DEX0259_153	Asn_Glycosylation 95-98;280-283;423-426;581-584; Camp_Phospho_Site 743-746;774-777; Ck2_Phospho_Site 47-50;145-148;189-192;218-221;552-555;657-660;665-668;702-705;715-718;762-765;896-899; Glycosaminoglycan 504-507;
30		Myristyl 320-325;355-360;386-391;469-474;807-812;814-819; Pkc_Phospho_Site 25-27;70-72;87-89;127-129;145-147;202-204;462-464;495-497;525-527;654-656;679-681;702-704;707-709;726-728;790-792;946-948; Tyr_Phospho_Site 318-325;
	DEX0259_155	Pkc_Phospho_Site 10-12;
35	DEX0259_156	Ck2_Phospho_Site 33-36;101-104; Myristyl 2-7;6-11;37-42; Pkc_Phospho_Site 27-29;33-35;40-42;
	DEX0259_157	Myristyl 52-57;
	DEX0259_158	Asn_Glycosylation 141-144; Camp_Phospho_Site 28-31; Ck2_Phospho_Site 104-107;106-109;143-146;184-187;207-210;
40		Glycosaminoglycan 12-15; Myristyl 37-42;166-171; Pkc_Phospho_Site 69-71;135-137;152-154;
	DEX0259_161	Ck2_Phospho_Site 45-48; Myristyl 16-21;
	DEX0259_162	Pkc_Phospho_Site 7-9;64-66; Tyr_Phospho_Site 37-44;
	DEX0259_164	Pkc_Phospho_Site 9-11;

	DEX0259_165	Pkc_Phospho_Site 20-22;52-54;
	DEX0259_167	Amidation 32-35; Myristyl 18-23;41-46; Pkc_Phospho_Site 25-27;
	DEX0259_168	Pkc_Phospho_Site 13-15;69-71;70-72;
	DEX0259_170	Asn_Glycosylation 44-47; Ck2_Phospho_Site 51-54;
5		Pkc_Phospho_Site 27-29;
	DEX0259_171	Pkc_Phospho_Site 3-5;
	DEX0259_172	Ck2_Phospho_Site 29-32; Pkc_Phospho_Site 24-26;
	DEX0259_173	Asn_Glycosylation 100-103; Ck2_Phospho_Site 27-30;
10		Glycosaminoglycan 70-73; Myristyl 13-18;71-76;75-80;96-101;
	DEX0259_174	Pkc_Phospho_Site 3-5; Tyr_Phospho_Site 84-91;
	DEX0259_175	Camp_Phospho_Site 8-11; Myristyl 2-7; Pkc_Phospho_Site 11-13;
	DEX0259_177	Myristyl 26-31; Pkc_Phospho_Site 3-5;
	DEX0259_178	Ck2_Phospho_Site 4-7; Myristyl 23-28;
15	DEX0259_180	Leucine_Zipper 18-39; Pkc_Phospho_Site 22-24;
		Amidation 5-8; Camp_Phospho_Site 49-52; Pkc_Phospho_Site
	DEX0259_182	52-54;
		Asn_Glycosylation 13-16; Myristyl 76-81;85-90;
	DEX0259_184	Pkc_Phospho_Site 12-14;51-53;90-92;
20	DEX0259_186	Myristyl 21-26; Pkc_Phospho_Site 28-30;45-47;
	DEX0259_188	Ck2_Phospho_Site 35-38; Pkc_Phospho_Site 39-41;
	DEX0259_189	Pkc_Phospho_Site 15-17;45-47;
	DEX0259_190	Ck2_Phospho_Site 10-13;16-19; Pkc_Phospho_Site 10-12;
	DEX0259_192	Ck2_Phospho_Site 23-26;
25	DEX0259_193	Pkc_Phospho_Site 43-45; Rgd 32-34;
		Ck2_Phospho_Site 71-74; Myristyl 14-19;26-31;39-44;
	DEX0259_194	Pkc_Phospho_Site 7-9;
		Amidation 86-89; Ck2_Phospho_Site 108-111; Myristyl 23-28;
	DEX0259_196	Pkc_Phospho_Site 84-86;
30	DEX0259_197	Ck2_Phospho_Site 5-8;
	DEX0259_198	Asn_Glycosylation 13-16; Myristyl 6-11;
		Asn_Glycosylation 279-282; Ck2_Phospho_Site 5-8;30-33;50-
	DEX0259_199	53;120-123;145-148;147-150;161-164;225-228; Myristyl 143-148;
35		Pkc_Phospho_Site 30-32;98-100; Tyr_Phospho_Site 250-256;
		Asn_Glycosylation 21-24; Myristyl 32-37; Pkc_Phospho_Site 22-
	DEX0259_200	24;33-35;
	DEX0259_201	Asn_Glycosylation 11-14;
	DEX0259_202	Myristyl 59-64; Pkc_Phospho_Site 49-51;
	DEX0259_203	Pkc_Phospho_Site 4-6;
40	DEX0259_204	Ck2_Phospho_Site 22-25; Pkc_Phospho_Site 45-47;
	DEX0259_205	Pkc_Phospho_Site 6-8;
	DEX0259_208	Asn_Glycosylation 10-13;
	DEX0259_209	Asn_Glycosylation 25-28;
	DEX0259_210	Myristyl 36-41; Pkc_Phospho_Site 11-13;39-41;
45	DEX0259_211	Myristyl 5-10;7-12;
		Ck2_Phospho_Site 39-42; Myristyl 25-30;29-34;
	DEX0259_215	Pkc_Phospho_Site 6-8;59-61;
	DEX0259_216	Tyr_Phospho_Site 22-29;
		Ck2_Phospho_Site 3-6; Pkc_Phospho_Site 15-17;

-135-

	DEX0259_217	Asn_Glycosylation 14-17; Ck2_Phospho_Site 40-43; Myristyl 6-11;37-42;
	DEX0259_218	Pkc_Phospho_Site 4-6;25-27;
	DEX0259_219	Myristyl 7-12;
5	DEX0259_220	Pkc_Phospho_Site 31-33;34-36;
	DEX0259_221	Ck2_Phospho_Site 47-50;82-85; Myristyl 37-42;
	DEX0259_222	Asn_Glycosylation 27-30; Ck2_Phospho_Site 7-10; Pkc_Phospho_Site 7-9;
	DEX0259_223	Asn_Glycosylation 54-57; Camp_Phospho_Site 36-39; Myristyl 20-25;41-46; Pkc_Phospho_Site 53-55;
10	DEX0259_224	Asn_Glycosylation 235-238;274-277;290-293;328-331;433-436;790-793; Camp_Phospho_Site 378-381; Ck2_Phospho_Site 5-8;47-50;58-61;192-195;381-384;424-427;513-516;577-580;603-606;768-771; Glycosaminoglycan 626-629; Myristyl 103-108;130-135;622-627;642-647; Phosphorylase 624-636; Pkc_Phospho_Site 14-16;47-49;58-60;240-242;245-247;320-322;388-390;439-441;703-705;760-762;765-767;768-770;792-794;
15		Tyr_Phospho_Site 350-356;675-683;676-683;724-732;
	DEX0259_225	Ck2_Phospho_Site 39-42; Myristyl 46-51; Tyr_Phospho_Site 23-30;
20	DEX0259_226	Tyr_Phospho_Site 6-13;
	DEX0259_229	Pkc_Phospho_Site 10-12;22-24;
	DEX0259_230	Myristyl 40-45;65-70;86-91; Pkc_Phospho_Site 8-10;79-81;
	DEX0259_232	Camp_Phospho_Site 5-8; Pkc_Phospho_Site 3-5;
25	DEX0259_234	Glycosaminoglycan 11-14; Myristyl 14-19;
	DEX0259_235	Asn_Glycosylation 7-10; Pkc_Phospho_Site 9-11;
	DEX0259_236	Leucine_Zipper 30-51;37-58; Pkc_Phospho_Site 36-38;55-57;
	DEX0259_237	Prokar_Lipoprotein 20-30;
	DEX0259_238	Tyr_Phospho_Site 23-29;
30	DEX0259_239	Prokar_Lipoprotein 8-18;
	DEX0259_240	Asn_Glycosylation 28-31; Ck2_Phospho_Site 37-40; Myristyl 25-30;

Example 6: Method of Determining Alterations in a Gene Corresponding to a

35 Polynucleotide

RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1

40 through 136. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al.*, *Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al.*, *Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then
5 cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al.*, *Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are
10 nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Mannheim), and FISH is performed as described in Johnson *et al.*, *Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium
15 iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical
20 Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological
25 **Sample**

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 µg/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells
30 are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results.

This page is not part of
the pamphlet!

WO 02-42776

2/3

Date: 30 may 2002

Destination: Agent

Address:

-137-

The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 μ l of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 μ l of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 μ g/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 μ g/kg/hour to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally,

-138-

- intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of
- 5 administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

- The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-
- 10 release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-
- 15 release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and
- 20 EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

- For parenteral administration, in one embodiment, the secreted polypeptide is
- 25 formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, I. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

- For example, the formulation preferably does not include oxidizing agents and
- 30 other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is

shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as
5 liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic
10 acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such
15 as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or
20 stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial
25 having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture
30 is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 $\mu\text{g/kg}$ of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and

-141-

separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove

-142-

detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

10 **Example 12: Method of Treatment Using Gene Therapy-*In Vivo***

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

15 The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) *Cardiovasc. Res.* 35 (3): 470-479, Chao J et al. (1997) *Pharmacol. Res.* 35 (6): 517-522, Wolff J. A. (1997) *Neuromuscul. Disord.* 7 (5): 314-318, Schwartz B. et al. (1996) *Gene Ther.* 3 (5): 405-411, Tsurumi Y. et al. (1996) *Circulation* 94 (12): 3281-3290 (incorporated herein by reference).

25 The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. et

30

al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 µg/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be

-144-

determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

30 Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea

pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

5 Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver et al., *Biotechnology (NY)* 11: 1263-1270 (1993); Wright et al., *Biotechnology (NY)* 9: 830-10 834 (1991); and Hoppe et al., U. S. Patent 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, *Mol Cell Biol.* 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a 15 gene gun (see, e. g., Ulmer et al., *Science* 259: 1745 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., *Cell* 57: 717-723 (1989); etc. For a review of such techniques, see Gordon, "Transgenic Animals," *Intl. Rev. Cytol.* 115: 171-229 (1989), which is incorporated by reference herein in its 20 entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., *Nature* 380: 64-66 (1996); Wilmut et al., *Nature* 385: 810813 (1997)).

25 The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a 30 particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of

-146-

interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503-512 (1987); Thompson et al., Cell 5: 313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, *supra*). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (i. e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or

-148-

alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids,

5 YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be
10 introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent 5,399,349; and Mulligan &
15 Wilson, U. S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the
20 cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function
25 of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually
30 and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments,

which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising
 - (a) a nucleic acid molecule comprising a nucleic acid sequence that encodes
5 an amino acid sequence of SEQ ID NO: 137 through 240;
 - (b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID
NO: 1 through 136;
 - (c) a nucleic acid molecule that selectively hybridizes to the nucleic acid
molecule of (a) or (b); or
 - 10 (d) a nucleic acid molecule having at least 60% sequence identity to the nucleic
acid molecule of (a) or (b).
2. The nucleic acid molecule according to claim 1, wherein the nucleic acid
molecule is a cDNA.
15
3. The nucleic acid molecule according to claim 1, wherein the nucleic acid
molecule is genomic DNA.
4. The nucleic acid molecule according to claim 1, wherein the nucleic acid
20 molecule is a mammalian nucleic acid molecule.
5. The nucleic acid molecule according to claim 4, wherein the nucleic acid
molecule is a human nucleic acid molecule.
- 25 6. A method for determining the presence of a prostate specific nucleic acid
(PSNA) in a sample, comprising the steps of:
 - (a) contacting the sample with the nucleic acid molecule according to claim 1
under conditions in which the nucleic acid molecule will selectively hybridize to a
prostate specific nucleic acid; and
 - 30 (b) detecting hybridization of the nucleic acid molecule to a PSNA in the
sample, wherein the detection of the hybridization indicates the presence of a PSNA in
the sample.

7. A vector comprising the nucleic acid molecule of claim 1.

8. A host cell comprising the vector according to claim 7.

5

9. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and (b) incubating the host cell under conditions in which the polypeptide is produced.

10

10. A polypeptide encoded by the nucleic acid molecule according to claim 1.

11. An isolated polypeptide selected from the group consisting of:

- 15 (a) a polypeptide comprising an amino acid sequence with at least 60% sequence identity to of SEQ ID NO: 137 through 240; or
(b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 136.

12. An antibody or fragment thereof that specifically binds to the polypeptide
20 according to claim 11.

13. A method for determining the presence of a prostate specific protein in a sample, comprising the steps of:

- 25 (a) contacting the sample with the antibody according to claim 12 under conditions in which the antibody will selectively bind to the prostate specific protein; and
(b) detecting binding of the antibody to a prostate specific protein in the sample, wherein the detection of binding indicates the presence of a prostate specific protein in the sample.

30 14. A method for diagnosing and monitoring the presence and metastases of prostate cancer in a patient, comprising the steps of:

-152-

(a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient; and

(b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the prostate specific marker in
5 a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of prostate cancer.

15. A kit for detecting a risk of cancer or presence of cancer in a patient, said
10 kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 6 in a sample of a patient.

16. A method of treating a patient with prostate cancer, comprising the step of administering a composition according to claim 12 to a patient in need thereof, wherein
15 said administration induces an immune response against the prostate cancer cell expressing the nucleic acid molecule or polypeptide.

17. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 11.

20

SEQUENCE LISTING

<110> Sun, Yongming
Recipon, Herve
Chen, Sei-Yu
Liu, Chenghua
diaDexus, Inc.

<120> Compositions and Methods Relating to Prostate Specific
Genes and Proteins

<130> DEX-0260

<140>

<141>

<150> 60/244,782

<151> 2000-11-01

<160> 240

<170> PatentIn Ver. 2.1

<210> 1

<211> 55

<212> DNA

<213> Homo sapiens

<400> 1

catttaattg gcttgtccca agcaactcat ctcaaatact aaatccaaaa gaaaa 55

<210> 2

<211> 380

<212> DNA

<213> Homo sapiens

<400> 2

caacgaacta gatgagaatg agaggaactg gaagaagggt cagatgaaag tgagatccat 60
acatccttct tcagcaactt gtgcctctgc tctgcacctc ccgcaattaa ctactgaaaa 120
aagaacacag cttcacaaaa gagattgtaa aatcaggaag tatatctaag tcacctccag 180
tagccgtaac tctaccttgt ccagtaaaag gtgatgcatt acaaaagttg acccttttaa 240
agaaaatttg cttgatttaa catatgaaga tgctcctttt tctctctgaa gcaagagaaa 300
caattgtctc tcccttctgc tctttaatct gtcattctgca catttttact gcgaattaag 360
aataaacggt ttaaaaatat 380

<210> 3

<211> 377

<212> DNA

<213> Homo sapiens

<400> 3

ctagaggatc caagctcgga attcggtctg agatacagga taaagactct gttaacatgg 60
tgactctgca gatgccttct gtggcagctc agacctcatt aactaacagt gcgttccaag 120
cagagagcaa agtagccatt gtgagccagc ctggttgccag aagttcagtc tcagcagata 180
gtagaatttg cacagaataa aaacccatag aatgcagtga atgtggtagt gctttcagtg 240
atcaattaca tcatatgtca caaaaaacac agaggaacaa actgatatat tcaaggtgga 300
aagcccttga ataaaacctt atggctaata agcatatact cagagaaaaa tagtatgaag 360

tggagactgg gaaattc

377

<210> 4
<211> 1044
<212> DNA
<213> Homo sapiens

<400> 4
ctagaggatc caagctcggga attcggctcg agatacagga taaagactct gttaacatgg 60
tgactctgca gatgccttct gtggcagctc agacctcatt aactaacagt gcgttccaag 120
cagagagcaa agtagccatt gtgagccagc ctggtgccag aagttcagtc tcagcagata 180
gtagaatttg cacagaataa aaaccatatg aatgcagtga atgtggtagt gctttcagtg 240
atcaattaca tcatatgtca caaaaaacac agaggaacaa actgatatat tcaaggtgga 300
aagcccttga ataaaacctt atggctaata agcatatact cagagaaaaa tagtatgaag 360
tggagactgg gaaattcttt tatgggaaga tagatcttct catcagtgac catagatcac 420
atcttcagtg agcttatagt tggtagaaat ataatgatca tggaaaagtc cttgttcaga 480
aacagtacgc cagtaggtat caggggggtt acacaggaga gaaacttttg gaagaccttt 540
gaaggctatg aatgtggcag ggttgctagt ggtacattct gccttatcct cagaggggaat 600
catatagaaa taaaactatg aaaatgtaac tagaacatct tcatcaaaat atgaaagaac 660
acacgaagca aataagccct gtgaaaagga gtattttaga gatttcgac agaaatctaa 720
catcattata tggcagataa tatacaggat gtgtatttta ggacaatata ccttgaatca 780
ctagtgtgata tgtcaatgac taattaaaag gggttgtcag tgttacacat cattggttaa 840
atztatagca cagtgtacct cttccccctt ttttgataag agtcttctat tcccaaccaa 900
gatcggtata tgattagctc ttgtgtttct ttgattccaa atttcttcac ttgttatctc 960
agactactga agctcttcaa aaggaaaaat gtattttaatt taataatgta acacaacaag 1020
tttggtatgtg ttttaacttta taaa 1044

<210> 5
<211> 432
<212> DNA
<213> Homo sapiens

<400> 5
aaaaaaaaa aagagcaaca tctgtataaa gaaaagaaaa tatTTTTgat tatctcaaag 60
ctaacttttt ctgggtccct caggaactag ggtcttatgc tgggaatatg gatttatgtt 120
tctgtatgca tttatgtata tctttcatct gtataatgag tgtatgtatc tgttatctct 180
atataaactc ctctatttgg ttattttctt tttctttcct ttttttgggt ttttaacttt 240
tcagaaaatg aagcactaga cagggaactta ggggactggc ctaggggacag agagattccc 300
atTTTggcat ttttaccttc caattaaagg taataaaaaga gtagcattta gcaaagaggg 360
gatattttaag gatctgatac cagagagacc ctaaactttc atcttgggta ccccaacccc 420
ttctaggtga aa 432

<210> 6
<211> 573
<212> DNA
<213> Homo sapiens

<400> 6
aaaaaaaaa aagagcaaca tctgtataaa gaaaagaaaa tatTTTTgat tatctcaaag 60
ctaacttttt ctgggttccc tcaggaacta gggcttatag ctgggaatat ggatttatgt 120
ttctgtatgc atttatgtat atctttcatc tgtataatga gtgtatgtat ctgttatctc 180
tatataaact cctcctattt gttattttct ttttctttcc ttttttgggt tttttaactt 240
ttcagaaaaat gaagcactag acagggaactt aggggactgg cctaggggaca gagagattcc 300
cattttggca tttttacctt ccaataaagg taataaaaaga gtagcattta gcaaagaggg 360
gatattttaag gatctgatac cagagagacc ctaaactttc atcttgggta ccccaacccc 420
ttctaggtga aagatggtca agcatagtga ttaaaaagta tggtttattt cccagattca 480

3

taattttactg gccataatcc tgagcaagtc cattttatatt tctgggtgcct cagttatctca 540
 taacatgggg ctaatcataa tacctgagag gac 573

<210> 7
 <211> 581
 <212> DNA
 <213> Homo sapiens

<400> 7
 aaaaaaaaaa aagagcaaca tctggtataa agaaaagaaa atattttttgg attatctcaa 60
 agctaaacttt ttctgggttc cctcaggaac taggggtctta tgctggggaa tatggattta 120
 tgtattctgg tatgcattta tgtatatctt tcatctggta taatggagtg gtatgtatct 180
 gggtatctct atataaaactc ctctattgg ttatttcctt tttctttcct tttttggttt 240
 tttaactttc cagaaaatga agcactagac agggacttag gggactgggc ctaggggacag 300
 agagattccc attttggcat ttttaccttc caataagggt aataaaaagag tagcatttag 360
 caaagagggg atatttagga tctgatacca gagagaccct aaactttcat cttgggtacc 420
 ccaaccctt ctaggtgaaa gatggtcaag catagtgatt aaaaagtatg gtttatttcc 480
 cagattcata atttactggc cataatcctg agcaagtcca tttatatctc tgggtgcctca 540
 gttatctata acatggggct aatcataata cctgagagga c 581

<210> 8
 <211> 61
 <212> DNA
 <213> Homo sapiens

<400> 8
 ccactttaat ttcttgacag agttttcatg tattggtata gtttccaaag ttctctttgg 60
 t 61

<210> 9
 <211> 725
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (35)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (421)
 <223> a, c, g or t

<400> 9
 cccttcattgt accgggactc aacattccac tctcnggtta gacagattat ctaaatagaa 60
 aatcgacaaa aaaacattgg attttaaactg aactctagac caaatggacc taacagacat 120
 ttacagaaca gtttatctaa cacctacaga ctatacatc ttctcatcag catgtggaac 180
 attctccagg atagaccata tgttaagcca caaaacaagt ctcaacaaat ttttaaaaat 240
 tggaatcata caaagtatct tctcagacca caagagaata aaactagaaa tccataccaa 300
 gaggaacttt ggaaactata caaatatcatg gaagttaaac atgctcttga ataactattg 360
 ggtcaatgaa gaaattaaga tggaaattta aaaatttctt aaaacaatag aaatggaaac 420
 ncaacatacc aaaacatgtg ggatacagca agagctatgg caagagggaa tcttacagta 480
 ataatgcct acatcaaaaa agtggtagaa atatttttaa taaaaattt atcaatgcac 540
 ctcaaggaac tagaaaagca aaaacaaacc aatcccaaaa gtagcagaca gaaagaaata 600
 atgaagagca gagcagacca aatgaaaca gacaaaaaac aatacaagg gtcaatgaaa 660

tgaaaagttg tttcttcaaa aagataaaca aaattgataa cccactagca taattaacca 720
agaaa 725

<210> 10
<211> 664
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (450)..(526)
<223> a, c, g or t

<400> 10
tgatgcttaa gaagtaaatt taggttattt ttataattat accattattt aagtgtatta 60
aatgctaatt aaatttggtt tgcttgattt tgttcatatt gataaaactt ggcttaagtt 120
tattgcagaa agatagtcct tgtattacct tgataatata ctcagatttt atgtgtttca 180
tatacgattt tcttctatat ggtttggttc tgtgatgtaa gttattataa ctgggtatac 240
actttacttt tcattcttgt tattagttct ttcataaaat gtctacgtat atacattaga 300
gaatatcaaa aatagctttc agtttggtcac ttaaaattat atgtttctat tgcaaactct 360
tcccaattct gttactaaaa tataaccaaa tgtaagaat tcaagtattt tagaattact 420
ttctttgtga agtaaaattac atatccatgn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnaaaa acagtgtctt 540
tacataggaa atcaaattat ttaaaaaaat gatgtcattt gatgctttct gactaagaag 600
gccttagcag ttctcagttt gggtaaactt tagaccaagt ttcttccctg actctaggcc 660
cctg 664

<210> 11
<211> 726
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (447)..(523)
<223> a, c, g or t

<400> 11
tgatgcttaa gaagtaaatt taggttattt ttataattat accattattt aagtgtatta 60
aatgctaatt aaatttggtt tgcttgattt tgttcatatt gaaaaacttg gctaaagttt 120
attgcagaaa gatagccttg tattaccttg ataataact cagattttat gtgtttcata 180
tacgattttc ttctatatgg ttgtttcttg tgatgtaagt tattataatg ggtaaact 240
ttacttttca ttcttggtat tagttcttct ataaaatgtc tacgtatata cattagagaa 300
tatcaaaaat agctttcagt ttggcactta aatttatatg ttctatttgc aaatctttcc 360
caattctggt actaaaatat aaccaaagt taagaattca agtatttttag aattactttt 420
ctttgtaagt aaattacata tccatgnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnaaaaca gtgtctttac 540
ataggaaatc aaattattta aaaaaatgat gtcatttgat gctttctgac taagaaggcc 600
ttagcagttc tcagtttggg taaactttag acaagtttct tcttgactct aggccctgg 660
aatcccttct ttagatttac ttagaaaact tgttattgta attttttctg cctgtccttt 720
gggatg 726

<210> 12
<211> 867
<212> DNA
<213> Homo sapiens

<400> 12

```

cagcttctct gtctttctca ccaggctctc aggtatgtgc caagggtgcc atcaagggat 60
ccccacccac catgctttct agacttcagt catttctgaa ccaacttcac caatttctgt 120
gaactacttg gactattatt tctttccaat tttctcttaa atcaatcatg taccaactca 180
ggctgggttac tctctttcaa atacatatga aaggagccat cccattgaaa ttgtttactg 240
atgtgctgtg taaacgttgg tccacaaaag agaccaccca gatgggagga gaagccgacc 300
caggccacgc ccagagagag cagctgggga cctgggcccgg gattgggaag aaagtcgtcc 360
aaagagcaag gccaggagcc gcaactgtcg gtgggagtgg ggggctctgt ctctctgccc 420
tgccaccagg cctccctccc atgacggtcc acccctgccg gaaccacctc cgccctccca 480
ccccaccccc tgctccccct ggctcttacc acctgccctt ccctcctagt tctctctctc 540
ctacgaaggc ttctctctgc ttcttggaag cctctattac tggcagctgc cccgggcctt 600
cctgggggag aaggtaacaa ggatttgggg ggattatagg ggtgtcatag aatcccaaaa 660
tatcagaggg ccctttgtca aggcccttta tgacacaaat ggggagaccg aggtcacagg 720
tgggtagaga cttgcctggg gtctcatggt caggtcagag cacagctgag actgaactca 780
gctctgcccc tgtccttcag cgggacctga cccctgacct tggcgctga cccctgtcag 840
aggagctgtc tcgcctggga gggtttc 867

```

<210> 13

<211> 1300

<212> DNA

<213> Homo sapiens

<400> 13

```

tagcagaggt agttaccttg cgtgctcggg atactatgag ttccgcatca tcccgcata 60
gcacctaac atccgatat cccgctagct caggagtaca agaggcatgt ggtacaaggg 120
ccagagtgcc aggtatcagc acttgtgagc agacagtga gacgggtgg ctggctgggc 180
ccttggggag gtgaagcgtc acggacgtcc tttcctgtcc ctctctgtct ctctgtctgc 240
ccgctcttcg ggctcctggt ccccagggtg gacgaggctc agagccggat ggatgctgag 300
atctggcagc tctgtccag ctttgcctgc ctaccccaac cccctcccca ggggctctct 360
ccgcatcccc agcctgcagc tgctctgaga gcagcacctc ctcttcttc ttctctctcc 420
tcctcctctt cctcagcttc tctgtcttcc tcaccaggct ctgaggtatg tgccaagggtg 480
cccatcaagg gatccccacc caccatgctt tctagacttc agtcatttct gaaccaactt 540
caccaatttc tgtgaactac ttggactatt atttcttccc aattttctct taaatcaatc 600
atgtaccaac tcaggctggt tactctcttt caaatacata tgaaaggagc catcccattg 660
aaattgttta ctgatgtgct gtgtaaacgt tgggccaaa aagagaccca ccagatggga 720
ggagaagccg acccaggcca cggccagaga gacgagctgg ggacctgggc cgggattggg 780
aagaaagtcg tccaaagagc aaggccagga cccgcactgt cgggtgggag tggggggctc 840
tgtctctctg ccttgccacc aggcctccct cccatgacgg tccaccctg ccggaaccac 900
ctccgccttc ccacccccac cctgctccc ctggctctt accacctgcc ctctcctct 960
agttctctct ctctacgaa ggcttctctc tgccttctgg aagcctctat tactggcagc 1020
tgccccgggc ctctctgggg gacaaggtaa caaggatttg gggggattat aggggtgtca 1080
tagaatccca aaatatcaga gggcccttg tcaaggccct ttatgacaca aatggggaga 1140
ccgaggtcca cgggtggtag agacttgcct ggggtctcat gttcagggtc gagcacagct 1200
gagactgaac tcagctctgc cctgtcctt cagcgggacc tgaccctga ccctggcgcc 1260
tgaccctgtc cagaggagct gtctcgctg ggagggttc 1300

```

<210> 14

<211> 183

<212> DNA

<213> Homo sapiens

<400> 14

```

gttcccttct gaagtttcag gtaggtgtga atcttccggg aactgtccc acccgttaca 60
gggtggcagg attgttctc ctcatccac cccatcagca cgtgctacct catcagcatg 120
tgccacttgc acgtgccatg tgcaagagca tttgaggtct cagagaagat cccaaagaat 180
aga 183

```

<210> 15
 <211> 2721
 <212> DNA
 <213> Homo sapiens

<400> 15
 atgccatga aggactgcc gagcgttccc cagccagcct actccctggg cagggctcac 60
 gggaagccca cgctgaccog ggaagcagca tcttcagaa cagctcaggc ccctctgggt 120
 ccgggctctc ctcttccca gcccagtc cagggacttc cgctgccacc attagaggag 180
 acagttcaca gcagctctgc cccaaagctc cctggagagg agcaggacat aggacctca 240
 gcagcccagg aaatggcccc ttgggacctt caggaggctc gtgcgcggc ccccccagc 300
 ccgaccgccc ccgccaccac caccagcgc cggggcgggc ctgcgcggc tcgggcggc 360
 ctccgagtg agccaccaaa gaaggaagcg gaggctggcc ccagatctca aaggagtg 420
 acagggctct ccacactccc acataaatc accccaaca caacaggcca cactgtggaa 480
 ttggggccac gtagtggggg tgccctggtg gtacctagt gcaatcacac tgatgacgat 540
 gatgaggatg atgatggtaa aggtggcagc aacaatgat acagtgacaa ggatgacaa 600
 agtagtgagg aaccaatagg gagcgaactc atccattact gtgaggacgg caccaagcaa 660
 gctgttcatg agagattgcc ccatagacca gacacctccc atcaggccca gtgttctgaa 720
 tcaaatttca acatgaggtt tgggaaggac aaatatccaa actatagcaa taagaatgat 780
 gagaacaatg atgatgatga agatgatggt gacaactctg tcccaggccc tgcctggag 840
 gctgtgtgct caatgcctgt ttggcagagg aggcacaga acctggggga ccgagcccca 900
 gcacctcaga ctccggaaat gacctgcttg agccagtcct gtgcctgccc cctttgtgct 960
 gtcttggag ggaggcagaa gcaggatgac aatgagggca agttcaggga cactgtggga 1020
 gatgagaaaa ctgtgtgtgt gaaaactctc cagcctgca gaggtgcca catggggctt 1080
 aagatgtcct gcctgaaagg acaccaggca ttggatttag agccctccg ccccaaatc 1140
 caggatgatt tcattcaag atccttaact aattacatat gcaaagacc tatttccaa 1200
 taaggttccc ttctgaagtt tcaggtaggt gtgaatcttc cgggacactg tcccaccgg 1260
 tacagggtgg caggattgtt cctctcatt ccacccatc agcacgtgct acccatcag 1320
 catgtgccac ttgcacgtgc catgtgcaag agcatttgag gtctcagaga agatcccaa 1380
 gaatagacag cgccttgtt agcacctggg ctgacaggct tctttgggag agatgacaa 1440
 gaatagccat gccgggaact tgccgtgtgg cccctctccc tttcccacc tgtgatgtgc 1500
 agggccactg accccagggt tctccctgc tccagtgc atggacagca tggggtcatg 1560
 ggcgtacaca cagggtgctg taccaggggt cagtatttaa catacttgct ttacagatgg 1620
 gaacaggagg gctcaggggg acactctcaa aattacacag cttttaacag gtggcagaat 1680
 tggggttcag acccagatct gggttcaagt cactcatggt gtgattgccc cagttccttc 1740
 ccgcatctgg gccttgccat ctctctctcc gagtggacat ggagaggacg gggggccagc 1800
 agctggatgg ctgcagggat caagtcttct ctggggctgg cagtagaag agcatgtggc 1860
 tgggtggacg gcattgctgg ctctcacct ggcagtctcc tgccctctaa ccggtgtct 1920
 cttgttcccc tagtgccctc ggctagcatg acccgctga tgcgggtccc cacagcctct 1980
 ggttccagcg tcaattctct ggatggcacc cgcagccgct cccacaccag cgagggcacc 2040
 cgaagccgct cccacaccag cgagggcacc cgcagccgct cgcacaccag cgagggggcc 2100
 cacctggaca tcaccccaa ctccgggtgt gctgggaaca gcgccgggc caagtccatg 2160
 gagggtctct gctagggcgg ctgccagct gccgccccg gactctgac tctgtagtgg 2220
 cccctctct cccggccct tttccccc tgccctgcat actgcgcta actcggtatt 2280
 aatccaaagg ttattttgta agagttagct ctggtggaga caaatgaggt ctattacgtg 2340
 ggtgcccctt ccaaaggcgg ggtggcggg gaccaaagga aggaagcaag catctccgca 2400
 tcgcatctc ttccattaac cagtggcgg ttgccactct cctccctcc ctgagagaca 2460
 ccaaactgcc aaaaacaaga cgcgtagcag cacacacttc acaaagccaa gcctaggccg 2520
 ccctgagcat cctggttcaa acgggtgcct ggtcagaagg ccagccgcc acttcccgtt 2580
 tctctttaa ctgaggagaa gctgatccag ttccggaaa caaatcctt ttctcatttg 2640
 gggagggggg taatagtac atgcaggcac ctctttttaa caggcaaac aggaaggggg 2700
 aaaaggtggg attcatgtcg a 2721

<210> 16
 <211> 923
 <212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (558) .. (660)

<223> a, c, g or t

<400> 16

tgaattttga	gattacacat	ttcttcttaa	acagcattga	gaggggttga	gcagtgtgt	60
gtgactgctt	gggtgcgctt	ccttggaatg	atgtgaggtg	tcacacagca	gctctcagat	120
cccctccagg	aattacatga	tgcccagaga	gagcattctg	gacattat	aattcctgga	180
ggggatcaca	ctgattgttt	gaggttctta	agttttgaag	ccttttaagt	gcagagtcca	240
gttctgggtc	tgagtgttgt	aacttggacc	acactcccac	atgacagctc	tcagagtgt	300
tctgtttttt	cgtgttttgt	gtctttgccg	tgataaaaat	gtagcagtgt	gtgaccagtc	360
cccagtgtta	gaactgtggt	cagtgttcat	ggggctgtag	cactgccagc	ctgaaccatg	420
tgacatagat	gcacacactt	gcacggacta	acgttttcta	catgactttg	gaaattgcat	480
cccgtaaaat	gcatacaact	gaataatctt	tccaaataag	agtgactgtt	tttattaact	540
agacagtagt	ctcagacnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	600
nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	nnnnnnnnnn	660
cagtgtctgg	gctgtctggc	cgggggccact	ttttgagaat	cgctcaagga	ccagggtgtt	720
tgagcttgc	tggttggctg	actcctaaca	gtgactgtac	tgaaagggtt	tgaatggcac	780
gtgttcaaac	aaggcagttc	gcaagagtgg	taggaagaga	cggttaggat	tcgagtgtgt	840
cttctgtttc	gtaaaagggg	gtttgaggga	aggttgtgta	gctcttttaa	aaagagttaa	900
tggcggccgg	tgagtggtg	gat				923

<210> 17

<211> 1353

<212> DNA

<213> Homo sapiens

<400> 17

agagatggcc	cagtccctta	gtggtagggc	tcaccctctt	gtccgcttgg	tatccctgtg	60
ggagtctctg	tcgtggatgg	agacttctct	tcattggctgt	gagagaggtg	caggcctgag	120
gcaagaggag	tgacgtggct	tttgtaacga	gcttctcaca	actgttttcc	agcccatttc	180
tggagaggga	gttttaggag	ccccacaaat	gttcattcag	aaaactccct	ggaagatccc	240
aaggatttga	agactggact	tctggaagtt	tatggggact	tagcagcaaa	ttgcttacaa	300
tgaacacaaat	gaactcttat	ttgggggaaa	tggattttta	attgcctttg	gaatagtgat	360
cttattttta	aatgtttaca	gatttccaga	acaaaaatga	tagattatta	tctgctcttc	420
gatggatttt	gagattacac	atttcttctt	aaacagcatt	gagaggggtg	gagcagtgtc	480
gtgtgactgc	ttggtgtcgc	ttccttggaa	tgatgtgagg	tgtcacacag	cagctctcag	540
atcccccca	ggaattacat	gatgccaga	gagagcattc	tggacattat	ttaattcctg	600
gaggggatca	cactgattgt	ttgaggttct	taagttttga	agccttttaa	gtgcagagtc	660
cagtctctgt	gctgagtgtt	gtaaacttga	ccacactccc	acatgacagc	tctcagagtt	720
gttctgtttt	ttcgtgtttt	gtgtctttgc	cgtgataaaa	atgtagcagt	gtgtgaccag	780
tccccagtg	tagaactgtg	gtcagtgttc	atggggctgt	agcactgcca	gcctgaacca	840
tgtgacatag	atgcacacac	ttgcacggac	taacgttttc	tacatgactt	tggaaattgc	900
atcccgtaaa	atgcatacaa	ctgaataatc	tttccaaata	agagtgactg	tttttattaa	960
ctagacagta	gtctcagaca	tcttttggag	ggtttgttca	gacagagact	gtgaggcccc	1020
acccctgagt	ttccagttcc	tgagtccctg	atctcagaat	ttgcatttca	cacaagttcc	1080
cacagtgtct	gggctgtctg	cccggggcca	ctttttgaga	atcgctcaag	gaccaggggtg	1140
tttgagcttt	gctgtttggc	tgactcctaa	cagtgtactgt	actgaaaggt	tttgaatgcc	1200
acgtgttcaa	acaaggcagt	tcgcaagagt	ggttaggaaga	gacgttgagg	attcaggttg	1260
ttctttcgtt	tcgtaaaagg	gggtttgagg	gaagttgtg	tagctctttt	aaaaagagtt	1320
aatggcggcc	ggtgcagtgg	ctgatgcctg	tgg			1353

<210> 18

<211> 74

<212> DNA

<213> Homo sapiens

<400> 18

tgcgacaaaa ctgttgggag ggctgtgctt cctctggggg ctccaggag gatctatttg 60
 ttggcccttc cagt 74

<210> 19

<211> 160

<212> DNA

<213> Homo sapiens

<400> 19

tcctctgggg gctccaggga ggatctatct gttggccctt ccagtgtgtt ggggtgccagc 60
 gttccacact tgtggatgcy ccgcctcaac ctctgcccac cttcatgtgt ccatctcctt 120
 tgtgtctgcy tctttacctc ttcttcttgt ctgtgttgcc 160

<210> 20

<211> 746

<212> DNA

<213> Homo sapiens

<400> 20

aagcgggggg gaacctggtc tcaggaatcc gtgccggggc gtatccttat tcacccccgt 60
 gacattacac gctggggacc gaactcagtc cgagatgtag actaggcaat aacgctaate 120
 taatcccggg acgattaacc ccactcagcc cgccctggct taatgtccgc tgggtgtgtt 180
 tggagacctc gagcatttaa atgaagcgtt cacaacaag agcacaggtt gctttgggat 240
 cataaaagag caggagagaa aggccttctg gggctctcgg aagatggatg tcaacgcaga 300
 gttttcaagg ctgagtacgg gttagccagg agaaagaaga gaaggtggag agagcaacac 360
 ggcagatata aggaaggag tgctggcttg ggaactgcag gtgctgggta gccagagctg 420
 ggaaggagtg agcagcagca tgggtgtggc attctggtgt ggggaaacca gatcatgaaa 480
 tgccatttta aggagtttgg attttatcct gaaaacaatc agcaaatcaa gcatatgcaa 540
 atcaagaagt gatttgatcg aatctgatga gaaagccact tttgacaatg tggacaggag 600
 atggggcaga tataatacta tattgaggtt acggtggaag gctagatagg agactggtca 660
 tccttatggg ggataaccaa gtaatgctaa gaagtgaagc gaaccagttg aaagaaacgt 720
 gggcctaggg ttgccagata ctctgg 746

<210> 21

<211> 786

<212> DNA

<213> Homo sapiens

<400> 21

aagcgggggg gaacctggtc tcaggaatcc gtgccggggc gtatccttat tcacccccgt 60
 gacattacac gctggggacc gaactcagtc cgagatgtag actaggcaat aacgctaate 120
 taatcccggg acgattaacc ccactcagcc cgccctggct taatgtccgc tgggtgtgtt 180
 tggagacctc gagcatttaa atgaagcgtt cacaacaag agcacaggtt gctttgggat 240
 cataaaagag caggagagaa aggccttctg gggctctcgg aagatggatg tcaacgcaga 300
 gttttcaagg ctgagtacgg gttagccagg agaaagaaga gaaggtggag agagcaacac 360
 ggcagatata aggaaggag tgctggcttg ggaactgcag gtgctgggta gccagagctg 420
 ggaaggagtg agcagcagca tgggtgtggc attctggtgt ggggaaacca gatcatgaaa 480
 tgccatttta aggagtttgg attttatcct gaaaacaatc agcaaatcaa gcatatgcaa 540
 atcaagaagt gatttgatcg aatctgatga gaaagccact tttgacaatg tggacaggag 600
 atggggcaga tataatacta tattgaggtt acggtggaag gctagatagg agactggtca 660
 tccttatggg ggataaccaa gtaatgctaa gaagtgaagc gaaccagttg aaagaaacgt 720
 gggcctaggg ttgccagata ctctggttct tctttattct ttttcaagac ggggtctctc 780

tctgtc

786

<210> 22
<211> 391
<212> DNA
<213> Homo sapiens

<400> 22
agcagggtatt tagtaggcac tcaataagct gatgtctacc tgttttcttg cttctcatgg 60
taacagttgt ttattgtgta gtttcagcat tataagtttg ctcttagcct caaaggaatc 120
atthgtggga atccttcctt cttcatctta tcttttgtgt aaaataactt gaggatata 180
agcatacatt tctggggaaa tggaaatatt atttgctact gcaataataa ccaataaata 240
agtattttgc cagtcgtttt tggcatgtat gccttatatt tgtgtgcctt taaaaaaaaa 300
gcgtgtgtgt ttttaacaaa gcttactaat ttagatatgc caatttggtg acttcctgcc 360
agagtaagag tttttatggt ttccctctga g 391

<210> 23
<211> 566
<212> DNA
<213> Homo sapiens

<400> 23
agcagggtatt tagtaggcac tcaataagct gatgtctacc tgttttcttg cttctcatgg 60
taacagttgt ttattgtgta gtttcagcat tataagtttg ctcttagcct caaaggaatc 120
atthgtggga atccttcctt cttcatctta tcttttgtgt aaaataactt gaggatata 180
agcatacatt tctggggaaa tggaaatatt atttgctact gcaataataa ccaataaata 240
agtattttgc cagtcgtttt tggcatgtat gccttatatt tgtgtgcctt taaaaaaaaa 300
gcgtgtgtgt ttttaacaaa gcttactaat ttagatatgc ccatttggtg acttcctgcc 360
agagtaagag tttttatggt ttccctctga gacacacctt agtataactt tattttagtt 420
ttctgggcca ttgccatttg gcataacttc aaactagatg tgatggaccc acataaatac 480
ttttaagatt taaaattttt tttatggtgg ggggatatga aactttcaaa accagctgag 540
gtgtcagggtc tggaaatttt atttgg 566

<210> 24
<211> 123
<212> DNA
<213> Homo sapiens

<400> 24
cagatggaga gattcaagga acgtgggagg gggcacggag ctttcatgcc ctctccaggc 60
acgtaccctt ccaggaacct ccaaacggtt cagctgtctg gaagctctct gaaccttgtc 120
att 123

<210> 25
<211> 505
<212> DNA
<213> Homo sapiens

<400> 25
gaattcgcag ctggcagtga attgaagggt cacaggactc cctccttgag tttggttaac 60
ttgctagagc gggtcagaaa acacagaaat gctacttaca tttgctagtt tattaataaa 120
ggatatcata aaggatacag aagaacagcc agatggagag attcaaggaa cgtgggaggg 180
ggcacggagc tttcatgccc tctccaggca cgctaccctc caggaaacctc caaacgggtc 240
agctgtctgg aagctctctg aaccttgtca ttttggagtt ttatggaagc ttttttatgc 300
agtcattgatt gattacatca ccagcccttg gtgatcagct taaccttcag cttcgggtc 360

10

```

ctccctggag ggatgggtgt cccacccgc taatcctgcc ttggtctttc tggtgaccag 420
ccctttttct gaagccatct aagggccccc agccgccact catctcatta gcaaacaaaa 480
gataccacgg ctcaaaaaaa aaaaa 505

```

```

<210> 26
<211> 381
<212> DNA
<213> Homo sapiens

```

```

<400> 26
tctgtattca aaacaaaaat aaatgcagat aatacaataa tacagaagag tataaaatgc 60
aaagtgaag gtcttcctaa ccagtttcaa tcactagtgg caatcctgtc aacaatacct 120
ggttttattt cttcaaagga gaggcattaa ttttataaaa catactttta ttttctgaaa 180
ttttcacatc tttcccttct ttacaacccc atgggtgtcct ctttcttgta ttcttttttt 240
attccctga agcaattttt ttaaagatg catgaacagt aagcattctg atcctaacat 300
ctaaacatgt ctacagtgtg cctcacact tacctattag ttggctgggt atattcaggt 360
tgaaaatctt ttttttttct t 381

```

```

<210> 27
<211> 4893
<212> DNA
<213> Homo sapiens

```

```

<400> 27
atgaaggcag aaataaaggt gttctttgaa accaacgaga acaaagacac aacataccag 60
aatctctctgg acacattcaa agcagtgtgt agagggaaat ttatagcact aaatgccac 120
aagagaaagc aggaagatc caaaattgac accctaacat cacaattaaa agaactagaa 180
aagcaagagc aaacacattc aaaagctagc agaaggcaag aaataactaa aatcagagca 240
gaactgaagg aaatacagac acaaaaaacc cttcaaaaaa tcaatgaatc caggagctgg 300
ttttttgaaa ggatcaacaa aattgataga tcgctagcaa gactaataaa gaaaaaaga 360
gagaagaatc aaatagacac aataaaaaat gataaagggg atatcaccac tgatcccaca 420
gaaatacaaaa ctaccatcag agaatactac aaacacctct atgcaaataa actagaaaat 480
ctagaagaaa tggataaatt cctcgacaca tacactctcc caagactaaa ccaggaagaa 540
gttgaatctc tgaatagacc aataacagga gctgaaattg tggcaataat caatagttta 600
ccaacaaaaa agagtccagg accagatgga ttcacagccg aattctacca gagctgggca 660
gagacacaaac caaaaaaaga gaattttaga ccaatatcct tgatgaacat tgatgcaaaa 720
atcctcaata aaatactggc aaaacgaatc cagcagcaca tcaaaaagct tatccaccat 780
gatcaagtgg gcttcatccc tgggatgcaa ggctgggttca atatacgcaa atcaataaat 840
gtaacacagc atataaacag agccaaagac aaaaaccaca tgattatctc aatagatgca 900
gaaaaagcct ttgacaaaat tcaacaaccc ttcattgctaa aaactctcaa taaattaggt 960
attgatggga cgtatttcaa aataataaga gctatctatg acaacccccc agccaatatc 1020
atactgaatg ggcaaaaact ggaagcattc cctttgaaaa ctggcacaag acagggatgc 1080
cctctctcac cactctatt caacatagtg ttggaagttc tggccagggc aattaggcag 1140
gagaaggaaa taaagggtat tcaattagga aaagaggaag tcaaatgtc cctgtttgca 1200
gacaacatga ttgtatatct agaaaacccc attgtctcag cccaaaatct ccttaagctg 1260
ataagcaact tcagcaaatg ctcaggatac aaaatcaatg tacaaaaatc acaagcattc 1320
ttatacacca acaacagaca aacagagagc caaatcatga gtcaactccc attcacaatt 1380
gcttcaaaga gaataaaata cctaggaatc caacttacaa gggatgtgaa ggacctcttc 1440
aaggagaact acaaacact gctcaaggaa ataaaagagg atacaaacaa atggaaaaac 1500
attccatgct cagggggaag gaagaatcaa tatcgtgaaa atggccatac tgccaagaa 1560
ttggaaaaaa ctacttttaa gtctcatatg aacaaaaaaa gagctcacat tgccaagtca 1620
attctaacc aaaagaacaa agctggaggc atcacactac ctgacttcaa actatactac 1680
aaggctacag taaccaaaac agcatggtac tggtaacaaa acagagatat agatcaatgg 1740
aacagaacag agccctcaga aataacgcag catatctaca gctatctgat ctttgacaaa 1800
cctgagaaaa acaagcaatg gggaaaggat tccctattta ataaatggtg ctgggaaaaac 1860
tggctagcca tatgtagaaa gctgaaactg gatcccttcc ttacacctta taaaaaatg 1920
aattcaagat ggattaaaga tttaaacggt agacctaaaa ccataaaaaac cctagaagaa 1980

```

```

aacctaggca ttaccattca ggacataggc atggggcaagg acttcatgtc caaaacacca 2040
aaagcaatgg caacaaaaga caaaattgac aaatggggacc tagttaaact aaagagcttc 2100
tgcacagcaa aagaaactac catcagagtg aacaggcaac ctacaaaatg ggagaaaatt 2160
ttcgcaacct actcatctga caaagggcta atatccagaa tctacaatga actcaacaa 2220
atttacaaga aaaaaacaaa caaccccatc aaaaagtggg caaaggacat gaacagacac 2280
ttctcaaaag aagacattta tgcagccaaa aaacacatga aaaaatgctc atcatcactg 2340
gccatcagag aaatgcaaat caaaaccact atgagatacc atctcacacc agttagaatg 2400
gcaatcatta aaaagtcagg aaacaacagg tgctggagag gatgtggaga aacaggaacg 2460
cttttacact gttggtggga ctgtaaacta gctcaacct tgtggaagtc agtgtggcga 2520
ttcctcaggg atctagaact agaaatacca tttgaccag ccatccatt actgggtata 2580
tacccaaagg actataaatc atgctgctat aaagacacat gcacacgtat gtttattgcy 2640
gcattattca caatagcaaa gacttggaa caacccaaat gtccaacaat tatagactgg 2700
attaagaaaa tgtggcacat atacaccatg gaatactatg cagccataaa aaatgatgag 2760
ttcgtgtcct ttgtagggac atggatgaaa ttggaaatca tcattctcag taaactatca 2820
caagaacaaa aaaccacaca cgcataatc tcactcatag gtgggaattg aacaatgaga 2880
tcacacggac acgggggggg gaatatcaca ctctggggac tgttgtgggg tggggggagc 2940
ggggagggat agcattggga gatataccta atgctagatg atgagttagt ggggtgcagc 3000
caccagcatg gcacatgtat acatatgtaa ctaacctgca cagtgtgcac atgtacccta 3060
aaacttaaag tataataaaa aaaaaaaga aaaaaaaga ttttcaacct gaataatccc 3120
agccaactaa taggtaagtg tgagggcaca ctgtagacat gtttagatgt taggatcaga 3180
atgcttactg ttcatgcatc ttttaaaaaa attgcttcag ggaaataaaa aaagaataca 3240
agaaagagga caccatgggt tgtaaaagaa gggaaagatg tgaaaatttc agaaaaataa 3300
agtatgtttt ataaaattaa tgctctcctt ttgaagaaat aaaaccaggt attgttgaca 3360
ggattgccac tagtgattga aactggttag gaagaccttt cactttgcat tttatactct 3420
tctgtattat tgtattatct gcattttttt ttgttttgaa tacagattac ttgccgtctt 3480
cttttttttt ttgagacaga gtgtcagctc tagtcaccca ggctggagtg cagtgggtgca 3540
atcttggtct actgcaacct cagctcctg ccttagcctc ccaagcagct ggaactacag 3600
gtgtgtgcca tcatacctg ccaaaaacac cttttttctc agaattagaa aagaaaatcc 3660
taaattcaaa taaaacaaaa aatcctgaat agcaaaagca gtcctgagca aacaaaagaa 3720
atctgggggc atcacattag ctatctttta gtatactgc aaggctacag taatcaacac 3780
agtatgacaa ctggtataaa aatagacaca tataccaatg aaacagaata gagaacccag 3840
atataaagcc aaataactac aaccaactga tctttgacaa agcagactga aacatacaat 3900
gggggaaagg acacactatt cattcagtgg tgctggaaaa ttggatagcc atacacagaa 3960
gaatgaaact ggttccttat ctctcaccgt ctacaaaaat ttactcaaga tggattaaag 4020
acttaacctt ttccccattt gctctgagaa tactcgccag tggcacttgc ggctgcatca 4080
tttaccctaa agtaaatttg ccacaaaata gcacctatg attattatgt ttgcattgct 4140
ctagtatata aactttggaa acaaaagaca ttatcctatt tatagcactc catttttagt 4200
agcggtatct ccacttacaa aataaaatct cagccactga aaatgtcaaa tcttagaaaa 4260
cagcattcct acacatgatg ctaacattgt tcatgaaaag ttgttgacca aggattagtt 4320
tcatgaattc agtttttctg aaatagatta ttctgatgat tcagacaatt ctaatgttag 4380
ttctgttcag aaataactcc aagaacagtt ttatatattt attttcacat tgactatcag 4440
tcagatttgc ttcagcctca aagaacgtgt ttatgtaaaa ttaaatgagc gctggcagca 4500
agccacacct ttttttgaaa acagggaaaag ggtcaaatgt aagacctgaa accataacaa 4560
ttttagaaaa aaccttagga atatacttaa ccaaagaggt gaaagatctc tacacataga 4620
actgcaaaac actgatgaaa gaaaccatag atgatacaaa caaatagaaa acctcctatg 4680
ttcatgaatt aaaagaatca atattgtgaa aatgaccata ctgccccaaa caatctacag 4740
attcagtaga gttcctatca aaataccacc aatgcctgcc tacctttctt tctttctttc 4800
tttctttctt tctttctttc tttctttctt tctttctttc tttttctttc tttttctttc 4860
ctttttcttt ctttttttct ctctcttttt ctt 4893

```

<210> 28
 <211> 548
 <212> DNA
 <213> Homo sapiens

<400> 28
 tactgccttt ttcttttagt ctatccatta cagacatgaa gtgaaaggta tgtctctttt 60
 atttctaaag atgggtgaaag tgaaacccag aaagggtgaa tgcgaaggat tttctcagg 120

12

```

ggttcaagtt tgcagtttag ggagacccaa aggctagaac ccagatatcc cagattttat 180
tctttcttct cagtcattgt gcaatttctt ccacacttgt cccacactgg ggttgtgggt 240
gctgtctttg acttaaaaga tatctcttct taatgtcacg tgctacagga tattgcttga 300
aagacgtaac aggcaactgg atggctccat gcttgcttag tcagtcaaaa atccgtgaaa 360
cacttcttta ggaaaaaggg atagtgactg ggagtaatga tgaagcagtg aaataaagag 420
agggaagagg aagttgcctt gagaagtaat gatcataact aggtgttatt gggataggaa 480
atgtaattaa attaggggca aggggttgca gagctgagtt aactgcagtg attaagaaaa 540
aaaatcta
548

```

<210> 29
 <211> 584
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (14)..(15)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (31)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (54)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (64)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (71)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (353)
 <223> a, c, g or t

```

<400> 29
atccatatgc attnntattc ttattcatat ncgtgtggta ggaacagatg tttnaagtta 60
acanacacat ntcgtatgct ttgttttgaa tttagaagat gcagttgagc ttggaagcca 120
tattaatttg ctogaactat gttacgtttt atcatatata tcaattttct ggtcacacca 180
tacacttccc tcaacttttc ttaagtttgt gacattttca caccaaattg tcttacaact 240
tgatattttt atttattcta tttctctgtt tataactaca aagtacatac atatgtttat 300
tttctatgtg atacatttat atatgcatgt gtatatatgt ataaatcccc atnatagaac 360
tgtcactaat ttatgacacc ttatgggtatt atcatgttat agtttagtta cttttagaag 420
ttctctgctt accaagggaag agataatata taaatatcaa atgacaattg agttaagcca 480
gcttatgttt ttctgaaatt tcagttacaa atgtctgaag aaaagttgac tttgaaaatt 540
ttatgtggat gacattttgt ttctgtggac acttaaaacc ataa
584

```

<210> 30

<211> 738

<212> DNA

<213> Homo sapiens

<400> 30

```

gcagctatatt agcctgggac atgtcaaagt tgggtacatgg agacatcgac attgggtggct 60
gatgtcatca caatgcactg aggggttattt gtgctaata aggcaagacc cacggcaggg 120
ggccagggca ctgagtgagc aggtggatgc ctgagcctgg ccgccagagc atgggaccag 180
ccccctgggg aggcgaaggc agcaacagcc agtatgtggt tcatgaagtg ccgggaagac 240
tgctgcccct ccctcaggga ccaaaaactc ctggaggtcc acctccagac tcctctttgt 300
aaaaaccac attcccgtt cacacacgag tocctggcac ctccatgaca tcccatcact 360
ctgtcttctt tttaattgac agaaggagac cactgcccga gcaacttgccc cctaacaagc 420
cattaacctc cagttgtttg gctggaaaat gggaccttga tcccagagag aactgacttt 480
gtgtgaatac tcctgggtctg tcctccttat gcaggcagcg ctgggatcga agtggagacc 540
cccgagtggt gtgaacatcg cagcatccct gggaggggttc agcatgttta tcttgggctg 600
aaggttgtaa ttatgattta ggggaatgcc actttataat gactttttct tcgtttcatg 660
taaattaagt tcctagtgtg cgtgggaatg catatgtcac tattcctaag gggagttgga 720
gaattcatta gttactgt                                     738

```

<210> 31

<211> 496

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (488)

<223> a, c, g or t

<220>

<221> unsure

<222> (492)

<223> a, c, g or t

<400> 31

```

attatcttga ttgtggtgag aggttcttgg tataaatata tgcccgaacc aaattgtatg 60
ctttgaatat gtccatttat ttgcatatca gttatatattc aatgaagcta taaaagtaac 120
agacatccct ttccccgggc ccagggtctca ccgtcatggc atttggggaa actcagagga 180
actggtcctg gcatggggtc tcattttcct cagagcagat ggcataagct tcatgaagtg 240
gcagctgtgc ccctgcaccc agaccagagc ttggcgccac agtggaaacca cacacctcct 300
ttgccagagg ctgaatcact gttttatggc agagcagccg ccttgggcac tttcctcaac 360
tctcctgttt ttcacctgtg aactgggaca tcagtaatga tgggctcact agatcaaggg 420
agagaaagac tgtaaagaaa taaatgcata gaagcaatga ttaggttagga caggtgctgg 480
aaaagggnct ancaga                                     496

```

<210> 32

<211> 1583

<212> DNA

<213> Homo sapiens

<400> 32

```

attatcttga ttgtggtgag aggttcttgg tataaatata tgcccgaacc aaattgtatg 60
ctttgaatat gtccatttat ttgcatatca gttatatattc aatgaagcta taaaagtaac 120
agacatccct ttccccgggc ccagggtctca ccgtcatggc atttggggaa actcagagga 180
actggtcctg gcatggggtc tcattttcct cagagcagat ggcataagct tcatgaagtg 240
gcagctgtgc ccctgcaccc agaccagagc ttggcgccac agtggaaacca cacacctcct 300
ttgccagagg ctgaatcact gttttatggc agagcagccg ccttgggcac tttcctcaac 360

```

14

```

tctcctgttt ttcacctgtg aactgggaca tcagtaatga tgggctcact agatcaaggg 420
agagaaagac tgtaaagaaa taaatgcata gaagcaatga ttaggtagga caggtgctgg 480
aaagggaatc aacagatgac aaggtcacgg gagaggccct tcagatgctg gtctccaagg 540
gtctgcaggg gacgttgaa ctgaaagtgg acagcagcgg gccgtgcagc ctggcctgcc 600
gtgtaaagga cctggggctc gggctgagct tggtgaggcc ccagggggct ggaaggatgc 660
ctgtggccct cggagagcac agtgtcaggc aacggaatcc cagagtgcc ttgctgctgg 720
gatacctctt gccggagatc atctgctccc tgccccagag ggagcagccc agctctctgc 780
tctctgcaca cgggagcacg gacgttgcca ctgtttggag gagggcgccg caggtctacg 840
ccccgcctcg gcccaccgac cgctggccg tgccgcccct cgcacagcgg gagcgcttcc 900
accgcttcca gccacacat ccgtacctgc agcagagat cgacctgcca cccaccatct 960
cgctgtcaga cggggaggag cccccacct accaggccc ctgcaccctc caggttcggg 1020
accccagaca gcagctggaa ctgaaccggg agtcggtgcg cgcaccccca aacagaaaca 1080
tcttcgacag tgacctgatg gatagtgcga ggctggggcg cccctgcccc cccagcagta 1140
actcgggcat cagcgcacg tgctacggca gcggcggcg catggagggg ccgcccacca 1200
cctacagcga ggtcatcggc cactaccggg ggtcctcctt ccagcaccag cagagcagtg 1260
ggccgcctc cttgctggag gggacccggc tccaccacac acacatcgcg cccctagaga 1320
gcgcagccat ctggagcaaa gagaaggata aacagaaagg acaccctctc tagggtcccc 1380
agggggccag ggctggggct gcgtaggatga aaaggcagaa cactccgcgc ttcttagaag 1440
aggagtgaga ggaaggcggg gggcgcaaa tcgcatgcgt gtggccctcc cctcccacct 1500
cctgtgtgat aaatatattac atgtgatgtc tggctgaat gcacaagcta agagagcttg 1560
caaaaaaaaa aaaaaaagtt ttg 1583

```

<210> 33
<211> 284
<212> DNA
<213> Homo sapiens

```

<400> 33
gacctggcca atcagtcata taaaaaccta ggtgttctct gtagatatga caggaagaat 60
aaggaagata gacaggaagt tcttcctatt tctgtcttat cctgtgctgc tttttcctgt 120
catctctttc tcagggtctgt ctattctgga gcttgttgaa accattttgt ttggaagcaa 180
ttttaagaaa gaataatttt ttacataaat ctgtggtcca ggaatactct ggcaggtcta 240
aggcataggc attgttagtt gagaaggaaa gaaaatggat catg 284

```

<210> 34
<211> 429
<212> DNA
<213> Homo sapiens

```

<400> 34
taaaaataagg tgaaaaatac tctgaacagt actgccaaag gattaatacg cttcaggaaa 60
tagcagtaga agctacttga tgtgaaagaa tggaggaaaa aaggacagat cattagtga 120
ctgtggtacg acttcaagca gactaatatg tgtattttga atcctcggag gagagtggag 180
aggaagtatg tttcaagaag caatgaccaa aagtttcaaa ttgatgaaa actatatact 240
cagagattta aagagttgaa tgaactctag gcagaagaaa cacgaaacaa actacataaa 300
agcacaaatct tcaattccta caaactagta atagagaaga ttatgagaaa caattagagg 360
aattttaaaa gccacattaa gtacaggggg agcaaaaaata aaaatgacag cagaggcngg 420
gtgcggtgg 429

```

<210> 35
<211> 612
<212> DNA
<213> Homo sapiens

```

<400> 35
ccgccctttt ttttttcagt tacatttaat ttggggaata ggagataagt aacatttagg 60

```

15

```

gtccatattg gagcagcagc caggccaggt cagcaatgtg gctggggcac ccagttgccc 120
atgcctgccc ctctccgctc cttctctcat cttctctgca gtaaaagtca ggtgtttctc 180
aaactctaac ctgcacatga atcacacaga catctgttaa aatgcagact ctgagtcata 240
ggtctagagt tgggcctgag attctgcatt tccaacaagc ttctgagcaa taacagtgtc 300
tgggaccacg gaacataccc tgagcagtga ggtgctacag aacccccagc atctgtctct 360
aacaacccca aacagaatgg gcagagacag aggcattctag acttcaccag catatattca 420
aattctgact acaggggtatt ggtttaccac agaaccagag aagaatagca acacaaatcc 480
tatacgatat cttacgggtga tatctataga ccccaaatg gttaggaggc aagtacaaaa 540
ggctctgaaa ccccttacca atagccgata caatgtaact aaaactacta aatactctta 600
taatatctcg ga                                     612

```

<210> 36

<211> 856

<212> DNA

<213> Homo sapiens

<400> 36

```

cccaaatgca acaacagaat actcagaaag ttgaagccag taaagtgcct gagtatatta 60
agaaagctgc caaaaaagca gcagaattta atagcaactt aaaccgggaa cgcattggaag 120
aaagaagagc ttattttgac ttgcagacac atgttatcca ggtacctcaa gggaagtaca 180
aagttttgcc aacagagcga acaaaggcca gttcttacct agtggtctct atccccggac 240
agttccagga atattataag agtatttagt agtttttagt acattgtatc ggctattggt 300
aaggggtttc agagcctttt gtacttgcct cctaaccatt ttgggggtcta tagatatcac 360
cgtaagatat cgtataggat ttgtgttgct attcttctct ggttctgtgg taaaccaata 420
ccctgtagtc agaatttgaa tatatgctgg tgaagtctag atgcctctgt ctctgcccac 480
tctgtttggg tttgttagag acagatgctg ggggttctgt agcacctcac tgctcagggc 540
atgttccgtg gtccaagca ctgttattgc tcagaagctt gttggaaatg cagaatctca 600
ggcccaactc tagacctatg actcagagtc tgcattttta cagatgtctg tgtgattcat 660
gtgcagggtt aggtttgaga aacacctgac ttttactgca gagaagatga gagaaggagc 720
ggagaggggc aggcattggc aactgggtgc cccagccaca ttgctgacct ggcctggctg 780
ctgctcccat atggacccta aatgttactt atctcctatt ccccaaatta aatgtaactg 840
aaaaaaaaaa gggcgg                                     856

```

<210> 37

<211> 223

<212> DNA

<213> Homo sapiens

<400> 37

```

gctagcctcc caatagtgtc ggggtattact agtatgtgag tcaactgtggc tgggtgcctg 60
cctggggtga gatttaaatt ggccttgtaa gctaataaaa aatgaagtct attctgaggg 120
caatgtggag tcattgaaag gttcccagga aggaaaataa aaatccaaaa tcatgtttata 180
gaaaggtaac tcagccgggc accgtggctc atgcctgtgg tcc                                     223

```

<210> 38

<211> 256

<212> DNA

<213> Homo sapiens

<400> 38

```

gggtcaaatga atgctgttgt tgtaaaattt cagataatac aaagagttaa ccaataaaaag 60
aaaaagtcac tcataatctt accactatta acattttgat gtatctatct gtatgtatgg 120
ctattctttt ttggtaaaac atgatcctag cctatctaata aatttaataa ttggatttta 180
aaaatttaac cattatatta tgggtaacct tacatgtcaa taaacaattc cacattgtca 240
tgctttaaat ggctgc                                     256

```

<210> 39
 <211> 524
 <212> DNA
 <213> Homo sapiens

<400> 39
 catggctccc aagtgccgca gggtcacctgt tttcacagtc ccatcctccc acgtttctct 60
 tcagatggct tcatagagcc cagagctcct ctatacaaag tgtgatcatt cccagtggat 120
 ttcttcgctc catagcttta tcattggaga tctggttgat cctgacgtag cgctcaagaa 180
 agcactaaat ctgaaacgtt taaaaaccaa ttcacgtctc ctgagaacga tgttgataaa 240
 cacaattttt ttctttcctt ttgatcccaa aagaagaaaa tcatgacaat attctttcat 300
 aaatccatta ttacactatt actatgacag gatattgtat gtgggaaata atgaagccat 360
 ttgccgtctc ttcccagtt tccttttagag tttctgtgct gagcaaacct ccctgcgaag 420
 ttaatcagat gctggacttc ttccctcaat cacaccagtt gccagggag agagacactt 480
 acaggacact cccttctgcc tattcaagta gtgccccttc tact 524

<210> 40
 <211> 536
 <212> DNA
 <213> Homo sapiens

<400> 40
 gctggacgag ggcattggctc ccaagtgccg cagggtccct gttttcacag tccatcctc 60
 ccacgtttct cttcagatgg cttcatagag cccagagctc ctctatacaa agtgtgatca 120
 ttcccagtgg atttcttcgc tccatagctt tatcattgga gatctggttg atcctgacgt 180
 agcgctcaag aaagcactaa atctgaaacg tttaaaaacc aattcacgtc tcttgagaac 240
 gatgttgat aacacaattt tttcttttcc ttttgatccc aaaagaagaa aatcatgaca 300
 atattctttc ataaatccat tattacacta ttactatgac aggatattgt atgtgggaaa 360
 taatgaagcc atttgccgctc tcttcccag tttcctttag agtttctgtg ctgagcaaac 420
 ctccctgcga agttaatcag atgctggact tcttccctca atcacaccag ttgcccaggg 480
 agagagacac ttacaggaca ctcccttctg cctattcaag tagtgcccct tctact 536

<210> 41
 <211> 379
 <212> DNA
 <213> Homo sapiens

<400> 41
 atttcaggag aagctcttgg ccgctggggt ctctggccn ccatgaactt caggaagtgg 60
 gtgccataac agctgcctga actacagaat ctgggcactg gtgtagctct gtatgccctc 120
 cgtgtcagat gctggagatg tcatttgcac tgccagagtt tgccaagggt gcacacagaa 180
 agcagattga aaagcaccct cttggaacat ctctccaatg ccttctactc acaaagttta 240
 acatcattaa cacgtgacaa agaagaacta tttaatgggc ccagatctat ttatgaagac 300
 aatcaagtgg gagtttggag tggataaccc aaatttggat aactggtgaa taataaaatg 360
 tattttatttc tgctggtgt 379

<210> 42
 <211> 1215
 <212> DNA
 <213> Homo sapiens

<400> 42
 tttgggaaat ggatcaaacc acacttttag taaatgttat cactctatag cataagaaat 60
 aattattttt tttttatata aaaggctata gtataaaata tatgtatagt aattaaatga 120
 acacttgtga acctaatagc catatgaaga aaataacatt tctaatatct ttggatgccc 180

```

catgtactaa tgacagttat gcttttgcac tttcttgaat tttatgttta tttatctttc 240
ctctgtcatt atttataaatt ttatcacaca tggctgtatc ctttacatgt tttggcatta 300
tgtatttttg aactttttgt aaagacaatc ataccatgtg taattttcag ggacttgatt 360
tttttcattg acttttaagg gttcaaatat attatcactg tggctgtagt ttgccatatt 420
ttgctgatat agagcattca ttcacatgag ggtaggattc aggggccatc aagacagaga 480
aaacatacag taatgtgaat agggaaaagt aatatgaaga attattaatt gttacagcat 540
tggaacaatg aaatattgtc tagtaatatg taaagagaag tctcaagaat atgtgatgag 600
cagatgtaag gaattgctct tgtctccatg gtgaatttgg agcagccaat gaagagtccc 660
ctcacattgt ggcctcgtct aaagttaaga agtcgctgta gtgttgccct tgaagaatct 720
gcttcaaatt gacacttcag aactccccag aaacttgtct tctgggcca tgtgtaaagc 780
tgtttatgaa gaaatgtcaa gccagagggg ctctactaca aatttggcaa aggacaattt 840
caggagaagc tcttgccgcg tgggttctcc tggccaccat gaacttcagg aagtgggtgc 900
catagcagca gcctgaacta cagaatctgg gcactggtgt agctctgtat gccctccgtg 960
tcagatgctg gagatgtcat ttgcatattg agagtttgcc aagggtgcac acagaaagca 1020
gattgaaaag caccctcttg gaacatctct ccaatgcctt ctactcaca agtttaacat 1080
cattaacacg tgacaaagaa gaactattta atgggccag atctatttat gaagacaatc 1140
aagtgggagt ttggagtgga taacccaaat ttggataact ggtgaataat aaaatgtatt 1200
tatttctgct ggtgt 1215

```

<210> 43
 <211> 754
 <212> DNA
 <213> Homo sapiens

```

<400> 43
ggggctcaga agctgtgttg tgtatgttct ttccaagaat cccacctgtc tgctttcaag 60
cacacacggc gctagaaatt tagcctagcc tgagtcctgg gatgagagaa gagctaaaca 120
aagagacccc aaccgtcccc ttggccccct gcccgcgct tttgcagttt gccaaccttc 180
tagctagaca gccccctaag tctcgtgtt gcgagtgaag gagaattttt ctatttcac 240
ttccattga ccgaagcaga aaaattgaac cgaatctacg ccccttgctc tgattcctgc 300
tagaggaaaa cagaaaatca tcccgcaggt ctctttcagt ccctggatgg cgagcgagc 360
cctgggaggc cacacttagt tctttattgt gaatctctcg ctactcaagt tcggtcggga 420
ccagggcctc ggatggcctc ggttgcccgt aagtacgca aagaagaggt gaatccaatc 480
gctggcctag aggatagtga tcagacaacc cgaggattac taaacaaggg gcggcggtgt 540
ccctgtctca tggggttggc gtggggcggg gggtaggcag caagatcctc caggctcctg 600
gatgcaaaga gtgagaaaga aagcgagca tctggcagcc tgcttataaa tgcagccttt 660
cggaagatga aacttgagc cttaggttgt cctcctttat atccatgttc caatcctctg 720
ggctttcctc gaaatgaata aaattgtgga aatg 754

```

<210> 44
 <211> 955
 <212> DNA
 <213> Homo sapiens

```

<400> 44
aaaggggccc aggagacgac ccctttcaga aagaacgtca cttcatcmaa ctcggtgag 60
ttattractg actccccgra aagktcaaca acgccttctc ttctcagccs caccgcgagg 120
agwtcaatcg ctttaccta ggtagcctct tgttcagggc tcagggactc ctgtcttaag 180
gtccttcttg ggctcagaag ctgtgttgtg tatgttcttt ccaagaatcc cacctgtctg 240
ctttcaagca cacacggcgc tagaaattta gcctagcctg agtcctggga tgagagaaga 300
gctaacaaca gagaccccaa cgtcccctt ggcctcctgc cccgcgctt tgcagtttgc 360
caaccttcta gctagacagc cccctaagtc tccgtgttgc gagtgaaaga gaatttttct 420
atttcatctt ccattgacc gaagcagaaa aattgaaccg aatctacgcc ccttgttctg 480
attcctgcta gaggaaaaca gaaaatcatc ccgcaggtct ctttcagtc ctggatggcg 540
agcgcagccc ctgggagggc acacttagtt ctttattgtg aatctctcgc tactcaagtt 600
cgttcgggac cagggcctcg gatggcctcg gttgccgta agtacgcgaa agaagaggtg 660
aatccaatcg ctggcctaga ggatagtgat cagacaacc gaggattact aaacaagggg 720

```

18

```

cggcgggtgtc cctgtctcat ggggttggcg tggggcgggg ggtaggcagc aagatcctcc 780
aggctcctgg atgcaaagag tgagaaagaa agcgagcat ctggcagcct gcttataaat 840
gcagcctttc ggaagatgaa acttgagtc ttaggttgtc ctcctttata tccatgttcc 900
aatcctctgg gctttcctcg aatgaataa aattgtggaa atgaaaaaaa aaaaa 955

```

<210> 45
 <211> 503
 <212> DNA
 <213> Homo sapiens

```

<400> 45
gatattgtatt aggcaaattc ccacccacc cccatttttg tctatagcac ttttagaatc 60
atcttgtcat ataattttaa aacagctggg atttagattg atactgcatt gaatttacct 120
atctatttgg gggagaatta tgccaaatga caatattgtg tcttgccatc taggaatatg 180
agattttccc atttttttcc agtctttttt atcaccttta gaaaagctat attgttttct 240
ttatatacca cttgcacgtt attagtggg ttaattccaa gatgcatcaa tattatagct 300
tttatgaatg gaatattttt cattgtattt tctaattgtt tgctggacta tatggaaatt 360
gatttttggc atgctgatat atccagcaaa aaactttact gaactctaatt gttttgtttc 420
tgagaggttt ctgatgttct gtttcttgca gggatgtctg aatcttccaa gtaaaaatgn 480
gtagactcct attttcctta gac 503

```

<210> 46
 <211> 206
 <212> DNA
 <213> Homo sapiens

```

<400> 46
ggctgacaaa ataactacct ttacctttat ttttgcattt tatactcaca accatatattt 60
ttttggcccc cttcccttta ttttaactca taactgatac ttaaagggtg tctgccttat 120
taaatacagc ctaggctgc aagtgcataa tatttaaaaa tttgcaactt tgacttttta 180
aaaatctggg cttgggatgg agcaac 206

```

<210> 47
 <211> 394
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (93)..(119)
 <223> a, c, g or t

```

<400> 47
attagtctta tgctgcttct gccattttca tttctgtaga cagaagagaa tttagaatgg 60
tttactgtct gtctagtggg ggacaaatta tannnnnnnn nnnnnnnnnn nnnnnnnnna 120
cagatgactg acaactgtta acttctcact atgtgccagg gactattgtg agttaactca 180
cttaatcctc atagccacc tttgaggtac ctataattat tctatagatg aagaagcaca 240
gacagagagg ttaattaaga gcaagtgttg gagttgaact cctgatattt ccccttttaa 300
gctgaagtcc atgacctgct tccaattcc tggcagccac acagttgctc tgcnatTTTT 360
cagtcttcta actttcaaca tagttacttt ttac 394

```

<210> 48
 <211> 135
 <212> DNA
 <213> Homo sapiens

<400> 48

```

gtcacataac atttcgggtg gccattaggg tgagctttta gatctaactg gccaaagggg 60
cttaagtaca atctttgatc agtaagtggc ttatgcctac ccagagacag cccctcagta 120
gccaggctgt gaaag                                     135

```

<210> 49

<211> 394

<212> DNA

<213> Homo sapiens

<400> 49

```

gtaaccatca ctagtatgtg aggettaaca cgacctctca tcatgactga acgacattca 60
gtactctgat ccaggagcac ctccatagga gtcaggcttt aaaataaaat cacactcatc 120
cctgacagtc tggcagaata tgtgcatgcc caagggtata ccctctctgg actgagtgc 180
gtatgaagat ccaactatta gtcctggctg aatgggaagc caaaatataa actccttcag 240
ctttgatagc aatctgcaag tcacataaca tttccgggtg ccattagggt gaggctttaag 300
atctaactgg ccaagggggc ttaagtacaa tctttgatca gtaagtgggt tatgcctacc 360
cagagacagc ccctcagtag ccaggctgtg aaag                                     394

```

<210> 50

<211> 730

<212> DNA

<213> Homo sapiens

<400> 50

```

tggtagaac atttctcctt tgttagcctt tagcatactt tataatttta caccttataa 60
acaggaacag tgcctatggg ttttaattagt gcttagttgt tttgttttgc tccttcattt 120
ttggctgaga aattaatgat atttggaat atctggagt cctttttctt gaaaagggtca 180
caaaccactg atttaaagag gatgactttg aaaattttagc tcacaatagt tgtgaaataa 240
atgtagtagt actttgtagc ttaaattccg gtaaaattat cactttgtca ttttgatctc 300
agaggagagc tattatttgt agcaactac aaatataaac taacgtggaa ttctgtgga 360
tcaaggcatg atacatattt atatgtgtgt gtgtgtattc ttttctgaac caatatgaca 420
ataagccatc tactctgaag tacagaggca gccatctatc attgacttat aaagctttga 480
ccccagttag agtgtgtgta agaaggaata ccttgaacac ttcagagtga agtcaccag 540
cttagctgag tggggggccac catgccttgc tcaaagcagg ttctccagtc agcaaacatc 600
agtcaaggca gaatctatag gcagtgccta ggaacacaga cgcatttcag atggtgagga 660
aaaagcaagt gaagcacaca atttgaatct tggaaatata ctttgaatcc atgggggtta 720
gaagacacag                                     730

```

<210> 51

<211> 953

<212> DNA

<213> Homo sapiens

<400> 51

```

cgggacaaca ggaccctatg aagggtgggc cacagcaaaa ggagagatga ttctagagca 60
tccagtcttc tagggcagca aaacaaccta aattttctaa gaggccacc agctgagggt 120
gccccgggg agggctgagg cgtcagggtg acggctccac tgccactca cctgcgacct 180
caaagccctc ctctccttg ggggtgctct gacagccacc tccagggcag gcgagtggcg 240
ctgggacaaa ggctggccc actgcgccc acccaagcag acggtccttc cccagacct 300
ggcgccaaac tggagtgaac gcccgaccac cgtgtctcac agggaaactg acaccagatg 360
cgaacttcca aatggatccc tccctgcaag tgtggagctg gcgctaccag gactgctct 420
ggccatgcgt ctaagacaca ggcagagggc gctgcccacc acgctggcga cggcctcaaa 480
gcccctgttc atgcctggga cagcgcccaa ggaccttgct catgcctggg acaggcccca 540
gggccccac tggctgcagt cagcagcggg cagggtggtg ggggaaggta tggacactcc 600

```

20

```

gtggggccgga gctgggagaa caaggcctat tattggacac ctggtggcca tggcaaccac 660
acaaggatgc ctgagactga aaatctgtgg gcttcaagga gctccagctc ttgcactggc 720
tgagtcacag tgactatata actcttactc ccacttttgg gacacttttt gagaggggaa 780
agggatccta tctaactaca cgggacagac atcgcccaag accgtcctga gcaagcctgg 840
acgctgtgac cctaacgatg aagggtgtccc gcagacaatg tccggggcag gcaccatgct 900
ctcccaacct accacagcca gatgtttttg taaagaacaa taaaaacgat tga 953

```

<210> 52
 <211> 527
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (224)..(365)
 <223> a, c, g or t

```

<400> 52
gttgttctct ggattaggag acagaagtta gagtcactat aacttttttt tttccctgg 60
aagttaatag ggggtatgta ttccttttagc aactgtatta tgtcttgagt atcaattgaa 120
atggccagtt taaggccgta atgtctaaat gggcaactat gctaacaata aaaaaagaac 180
attgaggtct attaatactg ttcacaaata tgggtgggttg tttnnnnnnn nnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnntcacc aatttacttt aacaatgcag agagaaagat ccattaacgt aagtgtttgg 420
atgagttgaa catgtgaaat atagattatt aaagtattga atgcatttta gatgtgggtt 480
atatatgggt tgtacttcat gaatattaag tctccacag caaactg 527

```

<210> 53
 <211> 406
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (308)
 <223> a, c, g or t

```

<400> 53
agagaatgat ggcacacagt aatgcctctt tctttatttt tgcagaaagt ttcattgagag 60
gggtgagaaac agcaggtatc caatattctg aaggatggca ttctgggggt gcctagggtta 120
ctcagcagga tgcattatca cattatgcct catattcttt tggagtaagt aaaaatgggc 180
aagatgtgag acatggaagt taagccttct gataagaaac ttgcatcatc atcactataa 240
tcaagaatgt gaaaagattt atttacacat cactttttta ttcatttatc cagtaatggt 300
agatgtgncc tgtctatgga actgtactag atgttgaagg aggtgtacct agaaatattc 360
agtctggttg aaaatatagg agatatacaa atgggcaggg tgttgt 406

```

<210> 54
 <211> 372
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (293)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (304)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (367)
 <223> a, c, g or t

<400> 54
 gttctttaac acatttgtat tatctttcag ttaaaagtat gtctttatgc ctacatattt 60
 caaagtaata tgagagagaa cattaactg tgttgtattg tgataaaatt cttggaatct 120
 taaacatcat aatacctcag gttatttggc cactgctctt gctagcaagg ctaagtagtt 180
 tcagtccttt agagctttat atttaaatga aggttaaaaa caaaaacggg atgggaagga 240
 acgtatcgcc taatacataa ttcttgtcat tagatgattt ttctgtaaa ggngctaata 300
 aggnatattc ctcggaattt attgtacatt atggattttg atatatactt agtaaagggt 360
 aagtaangga ct 372

<210> 55
 <211> 537
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (214)..(326)
 <223> a, c, g or t

<400> 55
 gcgtccgggc taaatgaaat atgaaataac catactattg aatactatgc atcagctaaa 60
 aatagcaaga gatctttggt gagtgaaaaa ataaattgct gattgatcat taaatataac 120
 actatgtttt taagaagcct cagaaaacag taatatatga tcctataggc ataaaattat 180
 ttatgatatc acacggagggt ctatagaatt tatnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
 nnnnnnnnnn nnnnnnnnnn nnnnnntagc aacatttgaa tgggtggccag tgtaatggag 360
 agtgcagatc tagaagaaca aacacaactg gtaacagagt tacctggggg aaggttgagt 420
 ttgggggatgg agggctacag aaactttaga gttctgcaga acttttaaca tttttacaat 480
 gagaatacat catatattat ctagctaatt taaaacaaat acattgttaa aatgaaa 537

<210> 56
 <211> 847
 <212> DNA
 <213> Homo sapiens

<400> 56
 caaaattaaa cttagacttt ttgaatttat tagctgtttt tgtgaagatt aatttttagaa 60
 agctaaaatt aaacactgaa agtaagttac tttattccat acgggtctctg tccagtttta 120
 gcactaaaat cagttcaagg atgccaatcc ctaattggcc aaatagcctt accattcttg 180
 ttttcttctc caaatttgtt tttttgctgg tcagataact tccaatctct aaaatattcc 240
 tgaaatgata aatttttatg atacagcata gaataatatg tatgtggaga cttgaaggag 300
 tcaaattctca atgagccttt tgtagggctt aacgattggt aaaagggggc caaaagggca 360
 ctaatttttg gaaagtgtat gtttgtttat ggtggtgaat gtgtagagag ggtgaaaagt 420
 aaaggaaaag tagaacaaga agaagaaaa ctgataggta tgacgatgag agagaaaaga 480
 aatgggaaga gagcgcaaga cgtgcagatt tagaaaaaag gttgagggaa acatattcaa 540
 aagggaaaaa gaaagcaggg ggaaaatata ttagaggtgt tgaaattagt aggcactcac 600

22

```

agaggtgcta atcgagagtt ctgttgggct cctgtcatgc tgctattaaa gagcattagc 660
agctaagaga tctaaattct agtcctagtt ctttgtgttg ccgtggagaa gtcagttaac 720
ttacatgagg ctcagggttc ttacctgtgt gtaaaatggg aacattgaac taggtgatct 780
ttaagatccc ttccgggtct aaaattgttt gacattatct tgggtgtcag taactgtgag 840
aaacaca
847

```

```

<210> 57
<211> 1448
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (1420)
<223> a, c, g or t

```

```

<400> 57
caaaattaaa cttagacttt ttgaatttat tagctgtttt tgtgaagatt aatttttagaa 60
agctaaaaatt aaacactgaa agtaagttac tttattccat acggtctctg tccagtttta 120
gcactaaaaa cagttcaagg atgccaatcc ctaattggcc aaatagcctt accattcttg 180
ttttcttctc caaatttggt tttttgctgg tcagataact tccaatctct aaaatattcc 240
tgaaatgata aattttttatg atacagcata gaataatatg tatgtggaga cttgaaggag 300
tcaaattctca atgagccttt tgtagggcct aacgattggt aaaagggggc caaaagggca 360
ctaatttttg gaaagtgtat gtttgtttat ggtggtgaat gtgtagagag ggtgaaaagt 420
aaaggaaaaa tagaacaaga agaaagaaaa ctgataggta tgacgatgag agagaaaaga 480
aatgggaaga gagcgcaaga cgtgcagatt tagaaaaaag gttgagggaa acatattcaa 540
aaggggaaaaa gaaagcaggg ggaaaataca ttagagggtg tgaaattagt aggcactcac 600
agaggtgcta atcgagagtt ctgttgggct cctgtcatgc tgctattaaa gagcattagc 660
agctaagaga tctaaattct agtcctagtt ctttgtgttg ccgtggagaa gtcagttaac 720
ttacatgagg ctcagggttc ttacctgtgt gtaaaatggg aacattgaac taggtgatct 780
ttaagatccc ttccgggtct aaaattgttt gacattatct tgggtgtcag taactgtgag 840
aaacacattc ctgaggaaaaa tttgcagcta tagttgactt caggacagca tgtttaggga 900
gtagaatgta agtccctga gggtaggggc cttttctggt gtgttcactg ccatatcccc 960
agcagctagc acaatgcgtg ttacatagta ggcattcatt aaatgtttgt tgaatgaatg 1020
atgtgaaaaa tatgttgatg gtttgttagg agcacaccta gaaagcctca aagaaaaatg 1080
gtgtgcttta gggagggaaa agacagattt cttctgaaga aatcttaagc aagctgattt 1140
ttaatcetta ttcttcttta ttttgtccca gattcaaaga aagtggcttc agctagtgc 1200
attctcatag tcacaaaact tacggtgact gtagacatac ataaaagtgt acatgtaatc 1260
taggccagtt ccctttaagt atcttacaga aaggcaggac caagcttagg tctccatgga 1320
atctgagtga aaagtatata catggaatat attagttata ttgaattaga ttgattggat 1380
taaaattcat tcagttgaga ggcacagtta gtctacaagn ctgagataca ggctgccaaa 1440
ttaagat
1448

```

```

<210> 58
<211> 354
<212> DNA
<213> Homo sapiens

```

```

<400> 58
acaaagatta ggacaagtat tccaggttct gacttacttc cttggagcct ctccttgaag 60
agctctgttt tctgaggacc gagtctaaaa actgaggccc tcagccactg gggacatgaa 120
atctcttga aaggaaaaat taagtcttgg gttgactagc aaaacctgac cttttcaagc 180
tctagctcta acatcttctt gtctctgagt tgctgctgaa agacaaaaat atgagagttt 240
gggacccatt tctactctc attctaatac agcagcagat attcattatt aatgaaatat 300
ataactatgt taatttaatt gatataggta ttgtttccag gatattcatt taaa
354

```

<210> 59
 <211> 586
 <212> DNA
 <213> Homo sapiens

<400> 59
 cactgcaaat gctactcgag gcagagagac ggaggaggtg gaatgtggcc tgtttccaca 60
 ttggggccctt cgggttttcca cagtgtcttt cactggcctt cttgaaatcc aggaaacaag 120
 agagctggaa aatattggtc tctgagttat agcacagggc agagaagggc agaaaatgca 180
 cctgaaagaa aacaggcaag tgacctatat accttctttt aggccttctc cctcttgtgt 240
 accgcatagc atattaagtg taaaattatt ataacactca ttgtatcacg tggctgtgtt 300
 ttgcttacat atccatctca acttttatct cttgctttcc ccagcaccag cactggcaca 360
 ttgcaatttt tgaacaaaag atttttgaac taatgaataa ataggtgatt agatttaatt 420
 caatttcaat gaatgtttat taggtcatta ttaggatatt gggtcagaat gttctagtgt 480
 attctacata catcacctcc ttcatagagt atcctgaaag gccacaaatt cactgcaca 540
 ttctttctcc taactgtcaa attttaccaa ttaaaaagta ttatca 586

<210> 60
 <211> 610
 <212> DNA
 <213> Homo sapiens

<400> 60
 gtgtggagga gacgcagcag ctaccactgc aaatgctact cgaggcagag agacggagga 60
 ggtggaatgt ggctgtttc cacattgggc ccttcggttt tccacagtgt ctttactgg 120
 ccttcttgaa atccaggaaa caagagagct ggaaaatatt ggtctctgag ttatagcaca 180
 gggcagagaa gggcagaaaa tgcacctgaa agaaaacagg caagtgcact atataccttc 240
 ttttaggcct tctccctctt gtgtaccgca tagcatatta agtgtaaaat tattataaca 300
 ctcatgtat cacgtggctg tgttttgctt acatatccat ctcaactttt atctcttgct 360
 ttccccagca ccagcactgg cacattgcaa tttttgaaca aaagattttt gaactaatga 420
 ataaataggt gattagattt aattcaattt caatgaatgt ttattagggtc attattagga 480
 tattgggtca gaatgttcta gttgattcta catacatcac ctccctcata gagtatcctg 540
 aaaggcccac aattcactcg cacattcttt ctccctaactg tcaaatttta ccaattaaaa 600
 agtattatca 610

<210> 61
 <211> 595
 <212> DNA
 <213> Homo sapiens

<400> 61
 aggaaatcaa ttaattttct tgaaaactgg aacatgaaat aatcaaacat ttattctgcc 60
 ttccttatat gaactatact actgaatagc caaatagatg aggggaagta tctttttgta 120
 atagtattct aactaatcaa ttaaaaagtg aaaataattt ttcagttctt attaaatgga 180
 tggacattaa acatcagtag ctactaagat tgcaaagtca gtcaaacatt agctatggat 240
 gttatagatg tcccaaagga atcagtcctg aatttgattc agtctcctgg atctagctgc 300
 ctatgacagg aaataaagaa taacatgtng gattgcagca tgagtatgta atctgcaaaa 360
 tccagactat gggaagcttg tcaggtcaaa gggcccaggt tcttttaaagc agaacttgct 420
 aggaaatggg tggaggaagg accaatagat taagacattc aagaaatc caatttttta 480
 atggatgaga ctaaaaaact gtgttcaagg atgcacattt gagtgcacaa actctgaaaa 540
 gacccaagga agtgattact attaaagtca aaacaacagt tggttatggg aggag 595

<210> 62
 <211> 810
 <212> DNA
 <213> Homo sapiens

<220>
<221> unsure
<222> (329)
<223> a, c, g or t

<220>
<221> unsure
<222> (691)..(752)
<223> a, c, g or t

<400> 62
aggaaatcaa ttaattttct tgaaaactgg aacatgaaat aatcaaacat ttattctgcc 60
ttccttataat gaactatact actgaatagc caaatagatg aggggaagta tctttttgta 120
atagtattct aactaatcaa ttaaaaagtg aaaataattt ttcagttctt attaaatgga 180
tggacattaa acatcagtag ctactaagat tgcaaagtca gtcaaacatt agctatggat 240
gttatagatg tcccaaagga atcagtcctg aatttgattc agtctcctgg atctagctgc 300
ctatgacagg aaataaagaa taacatgtng gattgcagca tgagtatgta atctgcaaaa 360
tccagactat ggggaagcttg tcaggtcaaa gggcccaggt tctttaaagc agaacttgtc 420
aggaaatggg tggaggaagg accaatagat taagacattc aagaaatata caatttttta 480
atgatgaga ctaaaaaact gtgttcaagg atgcacattt gaggacaaa actctgaaaa 540
gaccaagga agtgattact attaaagtca aaacaacagt tggttatggt aggagggaaa 600
agtattgtat aggcattggg agtatcgac agttaaata actcattaag ctaagtatat 660
ttgtatttgt ttgctgtatc tgttttattt nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 720
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnggccgagg tgggctagat ctacctgtag 780
gtcaggtagt tcgagaccta gcctggccat 810

<210> 63
<211> 1215
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (778)
<223> a, c, g or t

<220>
<221> unsure
<222> (801)
<223> a, c, g or t

<400> 63
agcaaataca gtacacataa aacatgggca tttgttctgg aaagggtttt ctctgtctga 60
tattgcagat agtttcacag gtcacagaac cttaaaaagg atttaaaggg catgtcttgt 120
gtagcatttg ttcctttgaa aatgatgctc ctttccattt ttttagtaat tgaagaggat 180
agaaagggtt tctcattgct tacgtttcac tgaattctct gcagcccctt tcccacaga 240
tgtttcagcc aaacctgtat ggaggagggt gacatggcat ggcttgctgt ttaaacagt 300
tacggtattt tgtgcttccc ttttgagtgt gtccaagttg aacaaaagga gagcctctag 360
aacgcattgg aggggaaatt tgggaccagg accttttaca tgctggggga aactgacagg 420
actcagttag gaaagacttt tgtttgtgtt ttcttctctc tctttctctg cagagcgcat 480
gatctatata aacatgcttc ctggtcatac taaagaatct cagctagtgg tgatctacca 540
gtttctgtga ggattattac tgtattaatg ctttttgga ggtgttcatt cagttcagag 600
tgaatgcttt ggaagacatt gcacagcttg aatcatgggg catcagggat agcttgactt 660
ttcctgaagg atgtatggtg gccatagact agttggttgg aagcttgcat tctgtaagcc 720
tggatatcaa tgcacacatt aagccatgtt ttcctaacag aatgaacatt ttttacannn 780
nnnnnnnnnn nnnnnnnnnn ngctcagaac cttagaacag gatgatatca tcagaaagaa 840
taagggaag taggccagaa ttagaaaaca tcaagatcat tggaaaactg ctatacttgc 900

25

```

attgcttccct ccttggttca ttgtacaatg gccttaattc aggtgacatt gcaagtacct 960
ttgggtgccct ccagaaatta agcgcatttg gtattgtgtg tgcagcttgt ttttcttctg 1020
ttgcagcaga caaaattgtg acatattatt gctaaggaga ttgacaactc ataagaataa 1080
atattgtctg tgggcaagat ttttttgttt gtttccagag aacattatta atttcagatt 1140
atattaaaga cttacatggc aggagacttt cttctagata actaaaaaca ctgcgtagaa 1200
agttatacta tgttt                                     1215

```

<210> 64
 <211> 1841
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (774)..(797)
 <223> a, c, g or t

```

<400> 64
agcaaataca gtacacataa aacatgggca tttgttctgg aaagggcttt ctctgctga 60
tattgcagat agtttcacag gtcacagaac cttaaaaagg atttaaagg catgtcttgt 120
gtagcatttg ttcctttgaa aatgatgtct ctttccatt ttttagtaat tgaagaggat 180
agaaagggtt tctcattgct tacgtttcac tgaattctct gcagcccctt tttccacaga 240
tgtttcagcc aaacctgtat ggaggagggt gacatggcat ggcttgctgt ttaaacacgc 300
tacggtattt tgtgcttccc ttttgagtgt gtcaaggatga acaaaggag agcctctaga 360
acgcatggga gggaatttgg gacaggacct ttacatgct gggggaaact gacaggactc 420
agtgaggaaa gacttttgtt tgtgttttct tctctctctt tctctgcaga gcgcatgac 480
tatacaaca tgcttcctgg tcatactaaa gaatctcagc tagtggtgat ctaccagttt 540
ctgtgaggat tattactgta ttaatgcatt ttgggagggtg ttcatcagat tcagagtga 600
tgctttggaa gacattgcac agcttgaatc atggggcctc agggatagct tgacttttcc 660
tgaaggatgt atggtggcca tagactagtt gggttgaagc ttgcattctg taagcctgg 720
atcaaatgca cacattaagc catgttttcc tagcagaatg aacatttttt acannnnnnn 780
nnnnnnnnnn nnnnnnnngct cagaacctta gaacaggatg atatcatcag aaagaataag 840
ggaaagtagg ccagaattag aaaacatcaa gatcattgga aaactgctat acttgcattg 900
cttctctcctt ggttcattgt acaatggcct taattcaggt gacattgcaa gtaccttttg 960
tgccctccag aaattaagcg catttggtat tgtgtgtgca gcttgttttt tttctgttgc 1020
agcagacaaa attgtgacat attattgcta aggagattga caactcataa gaataaatat 1080
tgtctgtggg caagattttt ttgtttgttt ccagagaaca ttattaattt cagattatat 1140
taaagactta catggcagga gactttcttc tagataacta aaaacactgc gtagaaagt 1200
atactatgtt tggccgggag cgggtggtc tgcctgcaat cccaacactt tgggaggcca 1260
agacattatc gaggaatttt ctggctgatt tctgggtcag tgccacagca gatcaattgg 1320
atggtcagtc cacgtcctgt ctccaaaggc ccagttccag agcccttgt gtctttggac 1380
attttcctca agtagcgcta gctgcaatgg ttacattgcc catgaaggac ctacctcagc 1440
tctgtctgcc gctccttgaa ggtacttcta ggagtctcca agatggcttg tgtgaacacg 1500
tgtcagacca gggtatttga ggccaccgtg ctgtcacctt cctctgccaa gtccaggccc 1560
actgtgggga ccgctgtcca ggcttagaaa ctccgtctcc cacaatttct ccactaagat 1620
gtgaaaatgg aagactagca ggcaagcctg tgggaacat ctgcgtcact ggcatctggg 1680
aaaagcaacc acccagggca ggatgccacg ggacagggga gcataagcaa ctgaaaatga 1740
agcggccaca aggccagagc ttggctcaca ctcagaattc gccaccctac catctcctgc 1800
caggaatatt ccaagaatgt ggagtaacag gggacagcta g                                     1841

```

<210> 65
 <211> 257
 <212> DNA
 <213> Homo sapiens

```

<400> 65
catgcctggc cttccacatg aaatttaaag tcagcttctc aatttctatt gttttgggtc 60

```

26

```

taaaatagat gtaaggggttt taaagtgage aacaatctct aggagccaga tttttgagtt 120
ttctctccca aagctgcttt tcccctagtc ttctccatct tagtgaatgg caacttcact 180
cttcagatg ctcacaccaa acaccctgaa atcactcttg attctttctc ttatacccca 240
cattaaattc ctcagca 257

```

```

<210> 66
<211> 327
<212> DNA
<213> Homo sapiens

```

```

<400> 66
caggcagtga tgcgaggtga tctagaggat cccgataccc attatgtgcg tgatcatagg 60
catgagccac catgcctggc cttccacatg aaattttaaag tcagcttctc aatttctatt 120
gttttggttc taaaatagat gtaaggggttt taaagtgage aacaatctct aggagccaga 180
tttttgagtt ttctctccca aagctgcttt tcccctagtc ttctccatct tagtgaatgg 240
caacttcact cttccagatg ctcacaccaa acaccctgaa atcactcttg attctttctc 300
ttatacccca cattaaattc ctcagca 327

```

```

<210> 67
<211> 487
<212> DNA
<213> Homo sapiens

```

```

<400> 67
gtaagtgttt attattatta cttctcattg tagtctcctt tatgaaacgt gtgtgcatag 60
cctgtctgga ggatgacttt ttgtctttta aagagagaag ctgtactact tctactgtac 120
cagaaattca tctgagagca gggtactttc tcattgtaaa gtccatgcaa gccagataaa 180
cctatagggt agcacttcct taattagttt acaatttctg aggatagggt ggtgggagta 240
aactgcctct gagtgttcac ttctctggga actgtcccgt ctgttggtgt gtatcatatg 300
ttctagtga ttttttttca gttatgtcct ttcccacaaa gcagtttggg gtaaccactg 360
taatcccagt aagctatggg tggggctctat gtataggaat gtgcaccctg aaattcattc 420
acttattcag cacaatttta tttgagcatc tactaagtgt tagggcactc tctgtggtca 480
gatatat 487

```

```

<210> 68
<211> 1006
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (317)..(479)
<223> a, c, g or t

```

```

<400> 68
aacattttat aaataacaag aaagagtatg ctactttcaa caatatcatg tttaatatac 60
ataaaatata taagcatgta aaatatatgt aacatatata cttaaaatgc atatacatta 120
tatacattta actaagtaca aatataaatg tgcctaagag gtaagcttca aatggaattg 180
agggaaataa gttcaaat catttctcat atattcatca ttttatttgt tcatatgtta 240
tgtttttgtt gttgtgtatg ggagaggtag tgatttaggt tacttctttg tagtagagga 300
tggtagttaa aaatacnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
atataatgtg ttggatcagt gcttatgtgg aagcactagg taagtgttta ttattattac 540
ttctcattgt agtctccttt atgaaacgtg tgtgcatagc ctgtctggag gatgactttt 600
tgtcttttaa agagagaagc tgtactactt ctactgtacc agaaattcat ctgagagcag 660

```

27

```

gttactttct cattgtaaag tccatgcaag ccagataaac ctatagggtg gcacttcctt 720
aattagttta caatttctga ggatagggtg gtgggagtaa actgcctctg agtggtcact 780
tctctgggaa ctgtcccgtc tgttggtgtg tatcatatgt tctagtgcac tttttttcag 840
ttatgtcctt tcccacaaag cagtttggtg taaccactgt aatcccagta agctatgggt 900
ggggtctatg tataggaatg tgcacctga aattcattca cttattcagc acaattttat 960
ttgagcatct actaagtgtt agggcactct ctgtggtcag atatat 1006

```

<210> 69
 <211> 126
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (70)
 <223> a, c, g or t

```

<400> 69
cccccttact ttttataagt attgatagct cccctttcat gcctgaggta ttaatctgag 60
tcttctcttn tttttttctt ggtcagctca gctaaacagt tgccaatttg ttgatctttt 120
ccaaga

```

<210> 70
 <211> 448
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (364)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (377)
 <223> a, c, g or t

```

<400> 70
tttttttttt ggaaaagatc aacaaattgg caactgttta gctagactga ccaagaaaaa 60
aaaaagagaa gactcagatt aatacctcag gcatggaaaa gggagcgaga ctctgtctca 120
aaacaacaac aacaaaaaga tacaagcaaa acaaatcaag aaacgtatac aaaggattat 180
acaccatgac caagtgggat ttatcccagg aatacaagggt tggtttaata tttgaaaatc 240
aatcgatgaa acacacaaaa ttgagagaat aaagatgaga agcttaatgt agggtaaaat 300
gtctgaagct ctaagtgaag ctgttgataa gctgggggtt ctactcttgg aacgctagag 360
aggnagagac acttagntac ttagtaacag caaaaagccc ggccaaaaag tagaactcaa 420
gtgctttaga aactctgtgg gcaggggt
448

```

<210> 71
 <211> 91
 <212> DNA
 <213> Homo sapiens

```

<400> 71
ttcggtcga gtaggaaatt aggaattaag taactgcct tcatactgggt aatcttgata 60
tgttgaagga agtgacttgt tataagatag a
91

```

<210> 72
 <211> 401
 <212> DNA
 <213> Homo sapiens

<400> 72
 aacaacaaaa aaaatccatt tataaataaa aatattttta aaaacaaaga gcttgcgatg 60
 ggcctgcaga cactcagcta aagatgtctc ataggttgtc cttgcagcta agtggggcca 120
 tgagactagg cttaaccag tgggctgaga gttaaagtga tttttgccat tctgttttta 180
 ggaatggatg tgtctgcctg tggcagatta tatttttcaa agatgaccac aaaaatatct 240
 cctatctcat gtgtgattct acagtggggt ctatgtcccc tcttcttgaa tgtgtgtgca 300
 cttgtgactg ctttgactaa cagagtatgg ggtaggatgc catgtgactt ctgaggctgg 360
 gtcacggaaa gcaattgtta taagttaaatt tgcattgtccc c 401

<210> 73
 <211> 422
 <212> DNA
 <213> Homo sapiens

<400> 73
 acatatgtag gtttgttata taacaacaaa aaaaatccat ttataaataa aaatattttt 60
 aaaaacaaag agcttgcgat ggcctgcag acactcagct aaagatgtct cataggttgt 120
 ccttgagct aagtggggcc atgagactag gctttaacca gtgggctgag agttaaagtg 180
 atttttgcca ttctgttttt aggaatggat gtgtctgcct gtggcagatt atatttttca 240
 aagatgacca caaaaatata tcctatctca tgtgtgattc tacagtgggg tctatgtccc 300
 ctcttcttga atgtgtgtgc acttgtgact gctttgacta acagagtatg gggtaggatg 360
 ccatgtgact tctgaggctg ggtcacggaa agcaattgtt ataagttaaa ttgcatgtcc 420
 cc 422

<210> 74
 <211> 471
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (392)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (459)
 <223> a, c, g or t

<400> 74
 ctttgtgtct ctacaacgta aatgtgaaaa gttagctcag acatagagga aacattcatg 60
 cttctatttt aagtagaaat gcctatgtga tactcaaaaa ttcttatttt agttgtacat 120
 cagaaagttc tgtttcacca gatcatgttt acagatagag tatgaggcat tgatccatga 180
 gaggacttca ttcaactaac ctttactgag cacctactgt atgcaatgca ccatttccga 240
 tgctaaaaaca ctgcaaagag gcagacagaa atccctaccc tgatggaatt ggcgttctgt 300
 gacacctctc taagtgtgtg ccccttccc tagtgctgtg acttacaatt ctttttaaag 360
 ccattattat tctggagaac ccaaggattg cntctttctc agagctctaa tgtcaataac 420
 cctatcatte tttgtcatag actttgcgaa ctgagggant cacatttaatt g 471

<210> 75

<211> 214
<212> DNA
<213> Homo sapiens

<400> 75
ggggtactca atgttagcct acagctcaac tcttactcta ataggatctc tttcctcctt 60
ctcccctaaa tttttccac tgggtgaaga gagatctgga tgactaaacc tcccctcttg 120
acaccttgga gtttgtaag cagggtcccct ctctgtagct tccaaagcca tgaagaagg 180
gaaggaaggc caagacaggg gtagatagag gtgg 214

<210> 76
<211> 214
<212> DNA
<213> Homo sapiens

<400> 76
cctccattca ccatctacag aatggaagag acgctaattgt caccctggaa ggtgttttga 60
agggtaattgt gtgtaaaagg ccaacaagg cccacacag ttaaggactt aatcctgccc 120
ggccccggga gggcttccgg catcttgggg ttcccctcaa aggatggcct gggcaggact 180
tcttaaaaac aaacaggcgg ctgggcgcgg tggc 214

<210> 77
<211> 552
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (273)..(357)
<223> a, c, g or t

<400> 77
aaatgtccca aatatgcagg ggggtcaagga acattgctca gtggaatcaa tggggagcgt 60
gggtactacc ccagcctctg gtttgcccca tagcagcccc ttgggcaccg ctggggaccc 120
ccaggcctct gaggagcttg gtttgaaag cgctggaatg ctggaccaag ttccctctct 180
ggctccctga gaggggtct tctagcccca gtcttagggc aagaggagcc cgtcccctag 240
gagcctccag gccctggagc cagacatcgg gcnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnaag 360
agacgctaatt gtcaccctgg aaggtgtttt gaagggtaat gtgtgtaaag ggccaaacaa 420
ggccccgcac agttaaggac ttaatcctgc ccggccccgg gagggcttcc ggcattcttg 480
ggttcccctc aaaggatggc ctgggcagga cttcttaaaa acaaacaggc ggctgggcgc 540
ggtggctcac gc 552

<210> 78
<211> 452
<212> DNA
<213> Homo sapiens

<400> 78
catttttagaa actagcatgt ttctaattaa acctgtttac atcaactaaa atactaacat 60
gctttgtaaa aggcccttca gtattgaaag tgtagtgcat tgggtgaagc ataataaaa 120
cacacaaagc cccattttaga tattgtgagg ctttcagtat ttagaatctc agtagtgatg 180
agtttaaaag gctaaggatg atggcaaagc tgattccaac ttggggctaa attttatttg 240
tcctgcttct atggaattag acctgagagt cacctatggg aggtcacaaa tgccttttca 300
attttgattt gcttgcattt tctatacagg ctgtaacact gccgcataaa acactagggg 360
ctcttgccag aggggactgt acaagcagtc cacagatgtt ctcgaagaaa ctccctggaa 420

ctttactact cggttataca aagagccgtc aa

452

<210> 79
 <211> 747
 <212> DNA
 <213> Homo sapiens

<400> 79
 tttttttttt ttttgagaca gtgtctcgcc ctgtcgccca ggctggagtg cagtgggtaca 60
 atctaggctc actgcaaccc ccgcctccca ggttcaagtg attcttatgc ctcagcctcc 120
 cgagcagctg ggattacaga tgcccaccaa gacactcagc caatttttgt attttttagta 180
 gagatggggc ttcaccatgt ttgtcaggct ggtcttgaac tcctgacctc aagtgatctg 240
 ccaccctgg cctcccaagt gctgggatta caggcatgag ccaccacgcc tggccttgac 300
 ggctctttgt ataaccgagt agtaaagttc cagggagttt ctctgagaac atctgtggac 360
 tgcttgatca gtccctctg gcaagagccc ctagtgtttt atgcggcagt gttacagcct 420
 gtatagaaaa tgcaagcaaa tcaaaattga aaaggcattt gtgacctacc ataggtgact 480
 ctcagggtca attccataga agcaggacaa ataaaattta gcccgaagt ggaatcagct 540
 ttgccatcat ccttagcctt ttaaaactcat cactactgag attctaaata ctgaaaagcct 600
 cacaatatct aaatggggct ttgtgtgttt atattatgct taccacaatg cactacactt 660
 tcaatactga agggcctttt acaaagcatg ttagtatttt agttgatgta aacagggttta 720
 attagaaaca tgctagtctt taaaatg 747

<210> 80
 <211> 353
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (102)..(217)
 <223> a, c, g or t

<400> 80
 ctctggggaa agccctgttc tgtgggcttt ggccacttaa atctatatgt cttctgctgc 60
 tgtctcaagg cagtgatgca gccctgacta tccttctgcc cnnnnnnnnn nnnnnnnnnn 120
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnct ggtaaaatgt acagaagcat 240
 gcatttttga aacaagtaaa ggaagaagac ttaggcgctc tccactccaa ggnacacctg 300
 acctncccta tgtagctttc ccagcaaca acgaagccna gcattggggt ctt 353

<210> 81
 <211> 627
 <212> DNA
 <213> Homo sapiens

<400> 81
 gaaaagtgc tttgggtgcc ggagacatgg gccctcccta gtcctggctg tctctaactt 60
 gtgagtgcact caagctgtct ccggttccac tttctggaag actaatggct tggatgagat 120
 cgtggttttc agatctgtcc tagccacaga accctctctt caaacaagcc cttacctgca 180
 atctgaacat aaaatgctac ctgggtgggac tcacgagtga gttccggaca ggacaggaga 240
 cggctgtcac tgggctccag gatgtgggtg gcagtgcctg acttcccgt cctgctgtct 300
 gtgggagacg agcttcttgc actggggcct gatttcccag gctggcctct cagatcccgt 360
 ggcttcaagt tctcctggtc atgcagtgtc ctgggtcagc actgaattgt tcctaattg 420
 tttcctgtgt ggcagattcc ttagctctac agtgaactct aacagggtag gcttgaccgg 480
 cttctgtggg ttgcttgag tagttaggat gaaaattcag aacctgcctg ctgactgaaa 540
 tgggcgttca tgtcttagaa tgctcaccag attgcttgtt ctcttacaca tagtagaggt 600

caataaaacg gagtttgtgg gatgttt

627

<210> 82
<211> 476
<212> DNA
<213> Homo sapiens

<400> 82
tcaataacaa ctttattctc tcatggctct ggattttaag atgttcaaaa ttaagggtgtg 60
ggcaggggcca tgttctctct gacaccttta gagaagagtc cttcctggct tcttagccag 120
cattgcccct tggttgcctg cagtccttgg tgtttcttgg ctgtagcaac atgactccga 180
tccctgtctc ctatccacac atggccttct gccctatat atctttgtgt cttgcacaag 240
gccttcttag aaggatacta gttgttggtt ttaagggtg caccctaate caacccatgg 300
cactcaatca ttaacctaaa ttaacattct gacgaaggag tcctatttcc ataataaagg 360
tcaacactga ggttactggg ttgaataatg gatatatgga catgtgtcct ccaaccccaa 420
atactcaata catatgaaat atgtaactac tcaagaaaat atacacacaa cagatg 476

<210> 83
<211> 387
<212> DNA
<213> Homo sapiens

<400> 83
caaaacaaag cntcaaagtc tacactgtgg cctgtagggc ccggcctggg ctggcctggg 60
ctgtgacttt gggcctcgtc tttctcttct cccctcctgg gtctctagac tccagcaacg 120
ttggcctcct tgctgectct tggcatgcca agctctctcc tcctgcaga cttcattcct 180
gctgttccct ctgttctaga tgcttcatca ttcaagcttc agcaaaagatg ccttttccct 240
gggggtggcct cccagcctg agcaacagca gcctctgctg gtcactcctg ccatgtcact 300
ccactctgtc tttccatagt gtctgttggg actgcaagta tcctattttg tgtatttggg 360
cattgtcagc gtcttctcag tagcatg 387

<210> 84
<211> 4270
<212> DNA
<213> Homo sapiens

<400> 84
atgctactga gaagacgctg acaatgaaca aatacacaaa ataagatact tgcagtacca 60
acagacacta tggaaagaca gagggtgagtg acatggcaag gatgaccagc agaggctgct 120
gttgctcagg ctggggaggc caccccaagg aaaaggcatc tttgctgaag cttgaatgat 180
gaagcatcta gaacagaagg aacagcagga atgaagtctg cagggaggag agagcttggc 240
atgccaaagag gcagcaagga ggccaacgtt gctggagtct agagaccag gaggggagaa 300
gagaaagacg agggccaaag tcacagacca ggccagacca ggccggggcc tacaggccac 360
agtgtagact ttgatgcttt gttttgggtt ggttttgggt tgtttttttt gagacggggg 420
ttcactctgt caccaggct ggagtgcagt ggtgcaatct tggttcactg cagcctccgc 480
ctcctgctcg tgcctcagcc tccaagtaga taggactaca gtggcgctc accatgcccg 540
actaattttt gtatttttaa tagagacggg gtttcgccat gttggccagg ctgggtctga 600
actcctgacc tcagccacag ctgttgcaaa tccaacactg tcctccttag atgttaaacy 660
gattttatatt caaaaaatta ccgacagagg ggatgagttg caaaaagcct ttcagctgct 720
ggatactggt cagaacttga ctgtgtcaaa aagtgaactg agaagaatca tcacagactt 780
cctgatgccg ctcacacgag aacagtttca ggacgtgttg gctcaggtgc tgaatatctt 840
cagcagatct ggtggctctc tggtcacatt taggtctaaa cagggtccagt gcaggcgagt 900
tcagtttaaa cctcaagtgt ggggctggtt cagcacgccc ttctcagctc ctccaagccc 960
gagggccctg gagggtgctt tgccaaactt cagtctgaca aatgataaga gccttgagga 1020
cgtcgctctg gtgggtgtga gctcgctgcc aggttgctgc ttcccggtgc tcacgctaga 1080
gatccatatt ggttgcctct ggggttttaa aaagcgcate gtgctcatgt cctccatttt 1140

accaaaggag	gttttcaaaa	acattaagac	tgttatgaaa	gcctttgagc	tcattgatgt	1200
taacaagact	ggactgggtc	gaccgcagga	gctaagaagg	gttctggaga	ccttctgtat	1260
gaagttaaga	gacgaggaat	acgaaaagtt	ttcgaaacac	tacaacatcc	acaaggatac	1320
tgcagtagat	tacaacgtgt	ttttgaagaa	tctcagcata	aataacgact	tgaaccttag	1380
atattgtatg	ggaaatcaag	aggtctcggt	ggagaatcaa	caagccaaaa	attccaaaaa	1440
ggaacgtttg	ctaggttctg	catcatctga	agatatctgg	agaaactact	ccttggatga	1500
aattgagagg	aacttttgtc	tacaactttc	gaagtcttat	gaaaagggtg	aaaaggccct	1560
cagtgcaggg	gacccctgta	aaggtggcta	cgtgtctttt	aattatctaa	agattgtcct	1620
cgacactttt	gtataccaaa	taccaagaag	aattttttatc	cagttaatga	aaagcatcaa	1680
ctctagaaat	gaatctcaca	aggaaaacat	catcaciaag	ttatttagac	acactgaaga	1740
tcactctgcg	tcactgaaga	aagcattact	gataatcaac	actatcaggg	acagcacgag	1800
aatgtcagct	ctcactgctt	ctagtcaaca	ctgtgctaga	gttctggaca	gtccaatcga	1860
gaaagaaaaa	gaaataaaa	gcatcaagat	tgaaaaggaa	gaaaaaccog	atggaccgat	1920
aacaagagaa	gaatttcgat	atattctaaa	ttgcatggct	gtaaaactaa	gcgattcaga	1980
attcaaagaa	ctaatgcaaa	tgcttgaccc	tggggacact	ggagtgggtc	acaccagcat	2040
gtttattgat	ctgattgaag	agaactgtag	ggtggaagaa	attgttcatg	atacaattac	2100
taggaaccta	caagcttttt	ataacatgct	acgctcatat	gaccttggag	acacagggcg	2160
cattggccga	aataatttca	agaaaatcat	gcacgtcttc	tgtccatttt	taacgaatgc	2220
acatttcata	aaactctgca	gtaagattca	ggacattggg	tcaggaagaa	tcctttacaa	2280
gaaacttttg	gcatgcatag	gaattgatgg	cccacccact	gtctctccag	ttcttgttcc	2340
aaaggatcag	ctgttaagtg	aacatttaca	aaaagatgaa	cagcagcagc	cagatctttc	2400
tgagagaacc	aagctcacgg	aggataaaac	caccctgacc	aagaagatga	ccacagaaga	2460
agtgttgaa	aaattcaaaa	agtgtataca	gcagcaggac	ccggcattca	aaaaacgatt	2520
tcttgacttc	agcaaggagg	ctaattgaaa	aattaacgtg	catgacttta	agaaggatct	2580
ggaagacact	gggatgcccc	tggacgatga	tcagtatgcc	ctgctgacca	ctaaaatagg	2640
cttcgagaag	gaagggatga	gctatcttga	ttttgcagca	ggatttgaag	aatctgtgct	2700
aaatgctggg	gatgaaagct	accagcccca	cccaggggag	ctcggctctaa	tcacagacac	2760
agatggtaaa	cagacaagtc	atattcatag	taattggctg	ttgtcgaagg	aactggagca	2820
tcctctggag	gcacaaggaa	aagatccctt	taaggaggac	tccttggagg	aggtggcccc	2880
cacactgaat	cgaatccctg	tgccctccaa	ccagcacgat	gggctgagtg	gaagtccttc	2940
agttgtggca	ggcttgtgga	tcccagacac	ggctgatctc	tatcacggtc	atcgaaacag	3000
aggcctggcc	ttcaggtggc	tccttgatgc	tgacagggat	ggcataatca	acatgcatga	3060
ccttcacaga	ctgctcctgc	atctactgct	taatctcaaa	gacgacgagt	ttgagcgctt	3120
ccttggcctt	cttggcttga	gacttagtgt	cactttaaat	tttcgggaat	ttcaaaattt	3180
gtgtgagaag	agaccatgga	gaacagatga	agcgcctcaa	agactcatta	gacaaaaaca	3240
gaaggttgcg	gattcagaac	tagcttgtga	gcaggctcat	cagtatcttg	ttaccaaaagc	3300
aaaaaacaga	tggtcagact	tgtctaagaa	ttttctagaa	accgataatg	agggcaatgg	3360
cattcttcga	cgccgggaca	taaagaacgc	actgtacggt	tttgatattc	ccctcacacc	3420
aagagagttt	gaaaagcttt	gggcaagata	cgacacagag	ggaaaagggc	acattactta	3480
ccaggaattt	ttacagaaat	tgggtattaa	ctattcgctt	gctgtccatc	ggccctgtgc	3540
agaggattat	ttcaacttca	tgggtcattt	tacaaaagcca	cagcagctac	aggaagagat	3600
gaaggagctg	cagcagagca	cagagaaggc	tgtggcagcc	agggataagc	ttatggaccg	3660
ccatcaagat	atcagcaaa	cattcaccaa	aactgatcaa	tcacaaaacca	actacatatc	3720
catatgcaag	atgcaggaag	tgctggaaga	atgtggatgt	tctcttaccg	aaggggagct	3780
gaccatctg	ctaaacagtt	ggggagtcag	ccggcatgat	aatgctatca	attacctcga	3840
cttcctgaga	gcagtggaga	acagcaagtc	aacaggagct	cagcccaagg	aaaaagaaga	3900
gagcatgcc	atcaattttg	caacgctgaa	tccacaggag	gctgtgagga	agatccagga	3960
agtagttgag	tcctcccagc	tggctttgtc	cacggcattt	tctgcattgg	ataaagagga	4020
tacaggattt	gtaaaggcta	cagaattcgg	acaagttctt	aaggatttct	gttacaaact	4080
aacggacaat	cagtatcatt	attttttgag	gaaactaaga	attcatctaa	ccccctatat	4140
aaattggaaa	tattttctcc	agaactttag	ctgttttcct	gaggagcgag	gcctgagtg	4200
cctcatgcac	gagtcacaag	gagcctctgc	tgatcctgga	attgctggac	caaggagaga	4260
acattcctaa						4270

<210> 85

<211> 468

<212> DNA

<213> Homo sapiens

<400> 85

```

agctaattct agagaaatta agcaaaagta tttttcttaa tatttttcct gaaacatttt 60
atgccatgct attgcttaga taatatataa tgtaatagca aaaagtcttt tccaaaatta 120
aataacttta aattatattt aaaaattctt taaaaccttt gtaacctatg taattcattg 180
tgaattgtta attattttta tgataggtag ttactttgat ttctctgaag tagcatgatc 240
ttgaggaagc cagagtcagt attcacacat gtoccaacag ggcttcttta gactatttgg 300
aataatattt ttgcctgcaa gacctgttaa ctcttcaagg ttttctgtat cttttcaaat 360
tggaaccact agaaaccacc agcttcttag ttatacctta gatatgctac accattttga 420
tgtagttggg tttgattact acaagattga tcccaactat taatacat 468

```

<210> 86

<211> 508

<212> DNA

<213> Homo sapiens

<400> 86

```

attttactca tcagaaggct tcactaggag atcacagagt gggaaataat atgttaatga 60
aatgttatat taagccaagc aaattatact ttaatatatta tagtatctca ttgaaaaaat 120
aaaactctat atgagtgtgt gttttgttta aataagcaac tacagaaaac atacatatga 180
acacacaaaa gagacactat gagattataa aagtgaagga atagtttatg agcctctgag 240
ctgcttaagc ttctaaaggc tgatagagta ggtaactaga aatgttgctt attatttcat 300
tctttaaaaa ctttttcaaa agtttagttg aagtctgcct ggaaactgtc tgggtgaagat 360
gatcaaggca atgaaaagga aactattaaa atctttaaaa tcttccttat tccaaatcca 420
cactgttgta ttgtcatatt ggcttcatta aaacaagaaa ttttattcat cagaagacct 480
cactaagaga cagagagact gaaaaagg 508

```

<210> 87

<211> 868

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (727)

<223> a, c, g or t

<400> 87

```

aacagacaag tccagagatc caaggaaagg ccagaagatt aagagtgtgg aagttttgag 60
gaaaagggaa tggagggagt ccattagaga aaaggataag ataaaaataa ggccaggccc 120
aagtcctaaa caacaccagc tattttgtca tggagtatag aaagggagca gccagtgaag 180
cagaacgaaa tcaggctctg gaggccttgt gcaagccatg agcaaagagg cggtcagccc 240
tgcaggtgat gcgggcaggt aagaaaagga cagaaggagc cggaccgctg gatgcaacaa 300
cttggagctc actggtgagc tcagtgatec acgtcagtg agacagagcc tgacgggtta 360
aaagtaaatg gaagggtgag atgagagaca tcacatatgc agacaattct cttagtgaact 420
aattccatat aatcagcaat tactaagaaa ttctaggcct tgtggctgca tggctgtgac 480
tccctgtggt ttggtctgat tacagctcct tcgaaagggt tcctggccag ctgtgaagcc 540
actcacagcc tcattgagac tgggctctcg cccgatgact cctgcagctc ctcaattgga 600
ctctaatac agagtaccgc tgctggcctt tttattttag ggagaatata acctccttac 660
tgatggctca cgaagccgca ctgccaggct acccaggtag accaacaagc accacttccg 720
aggcttnttc gctctgcca gcgtactggc aagccacctt ggttttcaca ttaccttta 780
attcacacca cgaggtgcc tcttaattcc ctgtgtatat tccactgcct tgaacgtag 840
cacattacgt ttcaattaaa aagaatcc 868

```

<210> 88

<211> 896

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (755)

<223> a, c, g or t

<400> 88

```

aatcgcagat gccagttaag aggccgcaaa cagacaagtc cagagatcca aggaaaggcc 60
agaagattaa gagtgtggaa gtttttagga aaaggggaatg gagggagtcc attagagaaa 120
aggataagat aaaatacagg ccaggcccaa gtccctaaaca acaccagta ttttgtcatg 180
gagtatagaa agggagcagc cagtgaagca gaacgaaatc aggcctctgga ggccttgtgc 240
aagccatgag caaagaggcg gtcagccctg caggtgatgc gggcaggtaa gaaaaggaca 300
gaagggaccg gaccgctgga tgcaacaact tggagctcac tggtgagctc agtgatccac 360
gtcagtggag acagagcctg acgggttaaa agtaaatgga aggtgaggat gagagacatc 420
acatatgcag acaattctct tagtgactaa ttccatataa tcagcaatta ctaagaaatt 480
ctaggccttg tggctgcatg gctgtgactc cctgtggttt ggtctgatta cagctcctct 540
gaaaggtttc ctggccagct gtgaagccac tcacagcctc attgagactg ggctctcgcc 600
cgatgactcc tgcagctcct caattggact ctaatcacag agtaccgctg ctggcctttt 660
tatttttaggg agaataaac ctccctactg atggctcacg aagccgcaact gccaggctac 720
ccagggtacac caacaagcac cacttccgag gcttnttcgc tctgccagc gtactggcaa 780
gccaccttg ttttcacatt acctttaaat tcacaccacg aggtgcctc ttaattccct 840
gtgtatattc cactgccttg aaacgtacca cattacgttt caattaaaaa gaatcc 896

```

<210> 89

<211> 229

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (113)

<223> a, c, g or t

<220>

<221> unsure

<222> (184)

<223> a, c, g or t

<220>

<221> unsure

<222> (202)

<223> a, c, g or t

<400> 89

```

caaaagtctc tcttccagct attttataat atattatacc tcctagaaac ataaatgtat 60
gctacaaaga aacatgtatc tatgtgtgta aacttaaaaa naattaatgg tancctttgg 120
gaagttttta ggagttgata tttatggtga agaaatatga agttcaggca ttctttgaat 180
ctancctcaa gttcttttta anatatattc aagttcccag cactttggg 229

```

<210> 90

<211> 234

<212> DNA

<213> Homo sapiens

<400> 90

35

cttatgaccc aaatttttag taggetgtta agaagatgcc atgtcttttt tccactagca 60
 ctttcaattt tctaaccaaa ataaaatgtt atgtcttctc caaggctgac cttttacctt 120
 ctagtctcag ttttggctca agccattacc agcactccca tcccccaacc ctaaaatgaa 180
 acttctcttc tgtttgttat ttctcttcct gacaatggat caacaaacat acat 234

<210> 91

<211> 326

<212> DNA

<213> Homo sapiens

<400> 91

ttcaagatca ctgagagcat aaagagatca ctcagttgac tgttatgtgg tgacttgaaa 60
 gtcttctttt ctaactttta tccttctttg atcttatgac ccaaattttt agtaggctgt 120
 taagaagatg ccatgtcttt ttccactag cactttcaat tttctaacca aaataaaatg 180
 ttatgtcttc tccaaggctg accttttacc ttctagtctc agttttggct caagccatta 240
 ccagcactcc catcccccaa ccctaaaatg aaacttctct tctgtttgtt atttctcttc 300
 ctgacaatgg atcaacaaac atacat 326

<210> 92

<211> 86

<212> DNA

<213> Homo sapiens

<400> 92

acaggcgtga ccaccggtgc ctggcccacg ctgtccttaa ggagacactt tgggtgcatac 60
 acagctgctc agcaaaacc gacttc 86

<210> 93

<211> 286

<212> DNA

<213> Homo sapiens

<400> 93

gagcaaatga taaaacaagc aggattaaac gttaactgtg tgtcagtcta agaggaacct 60
 ggctatcctt tgtaattcta ttgcagtctt tgtgtaaatt tcaggttact tccaaattta 120
 gaaaaaaatt aagtgaacac atatatgtac ccaaagttag acccattctg taacatgaaa 180
 atacaaggca aaaatatata taatacaact atgttaaaag accctttttt ctatcttacc 240
 taaaacttaa catctccaat gattatccat taataagctc ttttta 286

<210> 94

<211> 455

<212> DNA

<213> Homo sapiens

<400> 94

gataaaagta atgtattgat gtaaatttac tgcagttgat aactgtatca tggttgtgta 60
 aagtattaat aatatactca ttattgagaa atgcataattg aagtatttag aggtaaagaa 120
 gagtaaatgta tgaaattgaa atgattcaag aaaaatttgt gtatagaaag agcaaatgat 180
 aaaacaagca ggattaaacg ttaactgtgt gtcagtctaa gaggaacctg gctatccttt 240
 gtaattctat tgcagtcttt gtgtaaattt caggttactt ccaaatttag aaaaaatta 300
 agtgaacaca tatattgacc caaagttaga cccattctgt aacatgaaaa tacaaggcaa 360
 aaatatatat aatacaacta tgttaaaaga cccttttttc tatcttacct aaaacttaac 420
 atctccaatg attatccatt aataagctct tttta 455

<210> 95
<211> 158
<212> DNA
<213> Homo sapiens

<400> 95
ttttaaataa actttttgtt tgattacaac atgcatagtg tacaagtcac aagggtgccg 60
cttgatgaaa ttccacagtg accccagctg tgtaccacag atccagatca acaagcggga 120
ttacaggcgt gggccactgc gcctggcaaa ttgagcac 158

<210> 96
<211> 262
<212> DNA
<213> Homo sapiens

<400> 96
gtttttctgt gatgtgtacc taggaatgga agtgctgagc tctgtgtata cggcccttcc 60
tcatggttct aactactaga gctttatagt aagtcttggg atgtggtaag acatgccctt 120
cctccctctt ttcaaagtgt ccccaaaagg ctatacctag gtctttattc ttccttaaga 180
atttttcaac tgcattagat gttgccacct tatcttccaa agctgttggt gcagtttgtc 240
tttctcccag tgatatataa ga 262

<210> 97
<211> 87
<212> DNA
<213> Homo sapiens

<400> 97
atgagaaacg tacaaagaaa attttataat aagcgagttc agcaagggtg caagataaaa 60
gataagcata taaatagcag ttgtatt 87

<210> 98
<211> 230
<212> DNA
<213> Homo sapiens

<400> 98
gttcaggata aaagcttttag ggctgattct ccctcatggc acacattcac tgggcatctg 60
ctctttggca ggccctgtta taggtctggg actgcaaagc taaggcctgg tagtgtgact 120
acccggaata atcaggaaag gcatcaccaa ggcagcagta gctgtgctgt gatcaaagaa 180
tgcacagggc ttgtagctac aggagagaga gaacagtggc aattccaggc 230

<210> 99
<211> 144
<212> DNA
<213> Homo sapiens

<400> 99
gccttcattt ctagtggagc attcccaggc caaattaggt gaagggtctc atttcctagg 60
atttcttcac aggtggcatc cgtcctcaga tgggctacct aggactaggg atggctgcag 120
gtttcaagga gcgagtagtt gaat 144

<210> 100
<211> 469

<212> DNA

<213> Homo sapiens

<400> 100

```

gactaccaca caaggttatg catgttgtgc gatgttcagc tgtaggtggg gcgataactca 60
aatcgtagcc taggctgcta gtctttacat gcacagtgtg gtttagatgt gtgcttaatt 120
ctcacagaag ccctacgggg caggcattcc cgttttacag atgtggaaac aaactatgag 180
ggtaagaatt tggccagggt ttcacagcta ggatatggag ttgctgggat ctgaccgcag 240
tcctgtttcc ttcctaattc attggctgcc caccaggctg cccacgggg tgccccctggg 300
cagtcgctta tctatactat ctacctttac atacgttgat tggctggctg aggtgagtag 360
actaggactt gactggaaaa ttttacaac caagaaagca agggattctg ttcctcctac 420
ctcctagctt tctgtctcct agggaaagag aanattaca aagaagaaa 469

```

<210> 101

<211> 200

<212> DNA

<213> Homo sapiens

<400> 101

```

gggatgaatg gcagacttta actggatgct ttatttaggc ttttcgaaag caaaaaaagt 60
ttatacattg ttacagctgg gtgttgggtt acaggctgtt tggtatattc atgtattagt 120
tcctgttatt ttaacatttt aaatatttca taattgaaaa aggaaaaatt agactgggac 180
cagtttatag aaagctttaa
200

```

<210> 102

<211> 461

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (145)..(170)

<223> a, c, g or t

<220>

<221> unsure

<222> (435)

<223> a, c, g or t

<220>

<221> unsure

<222> (444)

<223> a, c, g or t

<400> 102

```

gggagaaaat agtgttgat gggagtaata tatttctatg tctctctggg atcaagctga 60
gatcaatatg tactgggatc aagtacatat tggagctga ttctgtaaat taagatatac 120
atggtaagcc ctagagtaac ccctnnnnnn nnnnnnnnnn nnnnnnnnnn acctcaaaaa 180
acatagttag ataaataatt taaattcttc attaggaaat atttacttaa tgcagaagaa 240
agcagtaagg gaggaataga agaacagaaa aatacatgag acacagtaaa ccaaaagtaa 300
aatgacagct ataaatccaa cttatatcaa acataacatt aaatgtgaat ggattaagga 360
atctgatcag aatgcagaga ttgtcagatg gattaaaata atncaataag gtccaactat 420
acactgtctg taggnacacac atgntagacg tgatgtttat a 461

```

<210> 103

<211> 319

<212> DNA

<213> Homo sapiens

<400> 103

```

gcttgccctta aggaacatga caaggatctg ttgtaagatc cacttcctaa agtgcttaaa 60
gaaagaaatg gaaatctcaa gctaaggctc cgagtcactg tgagggagac tttcccctc 120
cagtcctattc tgtagtaaca gaataaattt caaaataatt atttttccta attataaata 180
gaagtaatat cagctaattg tttaaagttt ggtaaatatt ttttaaatgt gaaaaaatc 240
ctctaatttc actcctaaaa ctcccttaac aatttgggta tctccagcct aggcaacaag 300
agtgaaaactc tgccacaca                                     319

```

<210> 104

<211> 563

<212> DNA

<213> Homo sapiens

<400> 104

```

tattaattaa gtactcgcta agtgctaacc accataccaa atgttggaaa tgtagtaatg 60
agtaggacat gtgtatatgg tccatacctg aaaggaagtt attctagtag gagaggtgat 120
ctatcaacac ataattacaa catgtgatat gagctatgaa cacttatgaa caaacaggg 180
gctgtgtaaa agaataaagg aacaaagatc tgtgtatagg agttttctgg aaaatgtttg 240
gattcggcag tcattttcaa aggcagaggg cattgatagc agtatcttaa catggaaaac 300
attaaaacta actagatatt agtattctat ttccaattca aaaataacca gaagatagtg 360
atgttgtttt gaatatagga tgtcaatctt tgtgttaatg tgttttgaaa aagcaagact 420
taattgaaaa tatacatcaa attataattt cagtgtatta aaaaactgcc tgttttaata 480
tgtcctttct ttgctgtaaa ttttggttaa aatctattgg agttacgtcc ttgtggtgaa 540
gtacacccta cccccaagag agc                                     563

```

<210> 105

<211> 1041

<212> DNA

<213> Homo sapiens

<220>

<221> unsure

<222> (140)..(229)

<223> a, c, g or t

<400> 105

```

ggtaagtcca tgatgttgat gttttgtaa catacccggt gtaggactat ggagcctatg 60
tctcagaaaa taaaacttga ataataatag aaaacaattt ttcataataa aaattatact 120
taagtataaa aatgtatacn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
tccgtacctg aaaggaagtt attctagtag gagaggtgat ctatcaacac ataattacaa 300
catgtgatat gagctgtgaa cacttatgaa caaacagggg gctgtgtaaa agaataaagg 360
aacaaagatc tatgtatagg agttttctgg aaaatgtttg gattcggcag tcattttcaa 420
aggcagaggg cattgatagc agtatcttaa catggaaaac attaaaacta actagatatt 480
agtattctat ttccaattca aaaataacca gaagatagtg atgttgtttt gaatatagga 540
tgtcaatctt tgtgttaata atgtgttttg aaaaagcaag acttaattga aaatatacat 600
caaattataa ttcagtgta ttaaaaaact gcctgtttta atatgtcctt tctttgctgt 660
aaattttggg taaaatctat tggagttatg tccttggtgg gaagtacacc ctaccccaaa 720
gagagcaaat gatgaataaa tcagtagatg ttccatgaat gcaatgttgg ctgagctggc 780
cacagtggag tgtgatcacc tggttatagg agaatagcca gcaggttata tttcataatt 840
atatttttcc ttaaatTTTT gcattaatat ttaatagcaa taattaaatg aattccagac 900
tgaatagaca attttattca ttgaataaac atgagaatt gcctactgag gcctgggctc 960
taggaattcc accaagaata aaaaagaca tgggtgtttg ccctcaaatt gcttagaatc 1020
tattcaggcc acttagtagc a                                     1041

```

<210> 106
<211> 451
<212> DNA
<213> Homo sapiens

<400> 106
tggcaaatat gtttttaaaa tggagaggtg tgcaggaagt gagccagcaa ggaaggagaa 60
tataagtcgt cttttttgca ggatgcaaaa ttgggtttat ttgcagactg atgtgttacc 120
ttctaaagga ctagccacaa cgtttgaccc tcaatctaag gtcaacactg ctatccattg 180
ctcacagacc agagtgcac tcccatgagg caaaagagca ggtgtgagaa gtgggtaagc 240
agtctgtata ttgggggtgt ggtggatggc ataggggata actcagtcta atgaaagaca 300
tcaatgtgcc attgggaaag gacagaggtt gccccctctt tccccagat agtcgcccag 360
cttataaatg catagatctg ggacagagaa taagggtcac ctagggtccc cctaatacaca 420
ggcgggacta ggacttttgg agatgtctca c 451

<210> 107
<211> 103
<212> DNA
<213> Homo sapiens

<400> 107
atcttgggcg gtctgaaatc tgagatactg tggaaagaac agaaagatcc tgtatctttc 60
ctataattgt tctactggaa gttgtcattt tacacaggag aca 103

<210> 108
<211> 979
<212> DNA
<213> Homo sapiens

<400> 108
agcggggggc ggcctgggac tcggggggcg ggtcagtcac ataaggctgt gccagcgct 60
tttgaagca gtaagtccag cccgaggcta aggaggtgtt aaccaccgaa ggggggtaga 120
atgtttttcc ccaccagagg aggcagcgac cacgtctcct ctatggaggc attcaagagc 180
cgtccagctg aagcagcatc actgtctgag ctcggaaggc acaatccaca taggtctgca 240
tgggtccacag agctgcatac ccacggggcc agcgggaggt gggcagctgc cgggctctct 300
tctgaagcag acaggatctc actctgttgc tgaggctgga tcacagctcc ctgcaacctt 360
gaactctccc tcaagcaatt ctccccactc tgccctccaa agcactagca ttataggcct 420
aagccaccac tcccatccac tgtagtgtaa actgtctcct tcaatgtttc caatagttgc 480
ggagcagatc agataagggt tcttcctgtc tgttgcttca agtttcattc tctctttaa 540
caatacaagg ttggcttcca tggttccttc ttaaagaatg ttgaagggtg gtcttcagat 600
tcatttagtg ttcgtggaac cccagggaaa gctgatgtaa aaacctcttt tttctcccat 660
atgtctcaaa aagttgtatt ttctgggtcc aagggatctg caagcctcct aaaggcattt 720
ccattgtcac taccaccagg tgtgaactgt aatctggcac gtatagttcc aagaactgtc 780
ataatagatg ctgaagaaac attgtgaagt taactcgctg ttaccaactg tgaagtcatt 840
agctagagga atcttgggcg gtctgaaatc tgagatactg tggaaagaac agaaagatcc 900
tgtatctttc ctataattgt tctactggaa gttgtcattt tacacaggag acattctgtt 960
ttattttatt tcttttgag 979

<210> 109
<211> 668
<212> DNA
<213> Homo sapiens

<220>

<221> unsure
<222> (583)
<223> a, c, g or t

<400> 109
tatcagcctt taaggtttat tgtcccacaa tggctgtgga gttaaaaaaa aaaattcagt 60
gagtttggat ataagattat tatttaatga ataatacataa cataggaaaa catatcaaaa 120
catagggaaa accaacataa atagtcttca aaagacacta gttcttggtta tattcacata 180
accacctttg tgaatgcagc acattaatac atctgtcata tagcacttta aaatggccaa 240
ctttttaagt gcttttatac tgtattctct ccacaatgat gtgactttcc aaaattttcc 300
actgaaaaag atgtaacctt gcaatgtggt ttagtatgga acttactttg cactgtatct 360
ggcggttgaa ttttgctttt attgtactgt ggacttgtga ctaaggcaaa taaaacttaa 420
gctcacttaa tttaaatata tcaaaaataac atttaggaaa aggtgcagtt tttctttgct 480
tcagaatggg tttttatcac aaaggaaatga gtgagacatt tatttgtgct gggacttctg 540
cacagtcatt gaatgctgtg agtgaatgtt aagtgaataa tcntgggtcaa ggggaaaacc 600
aaggtttcct ttccagggat aattcctacc caaattacct acctggaaaag gggaggaaatg 660
gccgagcc 668

<210> 110
<211> 1112
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> (17)
<223> a, c, g or t

<220>
<221> unsure
<222> (27)
<223> a, c, g or t

<220>
<221> unsure
<222> (59)
<223> a, c, g or t

<220>
<221> unsure
<222> (1027)
<223> a, c, g or t

<400> 110
aaaaatgccg gcccatngta ggggatncca gtcctatgcc ctttatgcct tcccagtcnc 60
aattaagacc ttgattgagc tgcagtagct ttaaaaaagga ttagaagagc tattgaatga 120
cttaatttat tagaagtttt taagtgcagc cttttctaata tattcaagtg catttatttt 180
tcatgaaaaa aggtagaatg atttggtctg acataaagta aatagtgttg atgcattaga 240
aattgtgtgt cttgattatg atttctgtac tttttgcatt agaagtataa tggacttgta 300
tttttaaata gttgaaacta gcaactgtgat catattaaat aatgcatttc tcagtttgga 360
cttcagatag ggattcattt gttgatattt tctttcttct ctcccctgct aacataaaca 420
cttttctgaa gcatatagtt atgatatcag cctttaaggt ttattgtccc acaatggctg 480
tggagtttaa aaaaaaaatt cagtgcagtt ggatataaga ttattattta atgaataatc 540
ataacatagg aaaacatata aaaacatagg gaaaaccaac ataaatagtc ttcaaaagac 600
actagttctt ggtatattca cataaccacc tttgtgaatg cagcacatta atacatctgt 660
catatagcac tttaaaatgg ccaacttttt aagtgccttt atactgtatt ctctccacaa 720
tgatgtgact ttccaaaatt ttccactgaa aaagatgtaa ccttgcaatg tggtttagta 780
tggaaacttac tttgcactgt atctggcggt tgaattttgc ttttattgta ctgtggactt 840

This page is not part of
the pamphlet!

WO 02-42776

3/3

Date: 30 may 2002

Destination: Agent

Address:

41

```

gtgactaagg caaataaaac ttaagctcac ttaattttaa tatctcaaaa taacatttag 900
gaaaagggtgc agtttttctt tgcttcagaa tgggttttta tcacaaagga atgagtgaga 960
catttatattg tgctgggact tctgcacagt cattgaatgc tgtgagtgaa tgttaagtga 1020
aaattcntgg tcaaggggaa aaccaagggt tctttccag ggataattcc taccctaaatt 1080
acctacctgg aaaggggagg aatggccgag cc 1112

```

```

<210> 111
<211> 1041
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> unsure
<222> (944)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (946)
<223> a, c, g or t

```

```

<220>
<221> unsure
<222> (976)
<223> a, c, g or t

```

```

<400> 111
gtcatcgtgc agttgtgatt taatttacac tcaatcacag ttcttgaata aattcttgaa 60
taaattgcaa aaccttgaga attacattat ttttatcaag tgctatcata tgtactaggc 120
tttttggtgca atttgacttc agatgttaat aaaacaaatc agaaaaaact aagggtgata 180
tttccaactg tggcttgctt catcatttgc gagactatgt catacatttc tacttttaga 240
cataacagaa gcagagagat tatactcaca gctaatatga ggtttttaa atcgtattat 300
atattcagcc tcagccagca tatcattttg gtggaggggt ggttacagat gattcaatat 360
tgtagtaatg ttgtcttctg aatttttttt cttagtattt tgtctgggtat gggatcatgt 420
agcttttttc tctttaactc gggtaattaa ggttcacaca gtaaagtcta tgcggtctaa 480
agctttaagg cggaggttgt tatctgttaa tgtgatggct ggtgccatca ggctctagac 540
gtttcttggt tcatgtcctg ggtttccctc ctggagaagt ccagtgaaaa agcatagctt 600
ttggagttgg tcagacttgg gttacagcgc cactctgcca ctactagct ggggggcttt 660
ggccaactac caaactctga tctccgtttc ctcacctata gagggtgagat gataaaacta 720
tattttattg attctaagat gcacagtttt tcaattttaa tctcttgga atcagaatgt 780
atcttaccgt tgggtgggtcc catataattg acagctgttt tcttttctna gaggtatgtg 840
caataatgat acatcttata atcagtggtg tcttagagtt gatgaattat ggtatttgcc 900
taaagaattt ttataaggat taaaatgtat tattcaagtg cttntntttc actatggcat 960
ataaagaggc caggggctgg aaaatgctca ggtgcatttc agttttgagc ttataaaact 1020
gggtagataa catgactagt g 1041

```

```

<210> 112
<211> 1380
<212> DNA
<213> Homo sapiens

```

```

<400> 112
tcgtgcgcgg taagaagctg cgcggtagcg cggtagagtg tgtttccatt gtagaaaacc 60
tggatcatgga attgcagatt gccccgccgc ccttgaaaat caagacatgg gcactgggat 120
atgttacagg tgtgggtcca cagagcacga aataaccaag tgtaaggcta aagtagaccc 180
ggctcttggc gaatttcctt ttgcaaatg ttttgtttgt ggagaaatgg ggcacctgtc 240
tagatcttgt cctgataatc ccaaaggact ctatgctgat ggttaagtact gttaccctca 300

```

42

```

tatagcagaa atggtgagtc atcgtgcagt tgtgatttaa tttacactca atcacagttc 360
ttgaataaat tcttgaataa attgcaaaac cttgagaatt acattatatt tatcaagtgc 420
tatcatatgt actaggcttt ttgtgcaatt tgacttcaga tgtaataaaa acaaatcaga 480
aaaaactaag gtgtatatatt ccaactgtgc ttgcttcac atttgtgaga ctatgtcata 540
catttctact tttagacata acagaagcag agagattata tctcaagcta atatgagggt 600
tttaaaatcg tattatata tccagcctcag ccagcatatc attttggtgg aggggtgggt 660
acagatgatt caatatgtga gtaatgtttg cttctgaatt tttttcttta gttatttgtc 720
tgggtatggga tcatgtagct tttttctctt taactcgggt aattaagggt cacacagtaa 780
agtctatgcg gtctaaagct ttaaggcgga ggttggtatc tgtaaatgtg atggctggtg 840
ccatcaggct ctagacgttt cttgtgtcat gtccctgggt tccctcctgg agaagtcag 900
tgaaaaagca tagcttttgg agttggtcag acttgggtta cagcgccagc actgccactc 960
actagctggg gggctttggc caactaccaa actctgatct ccgtttctc acctatagag 1020
tggagatgat aaaactatat tttattgatt ctaagatgca cagtttttca attttaatct 1080
cttggaaatc agaatgtatc ttaccgttgg tgggtcccat ataattgaca gctgtttttc 1140
tttctgagag gtatgtgcaa taatgataca tcttataatc agtgggtgtc tagagttgat 1200
gaattatggg atttgcttaa agaattttta taaggattaa aatgtattat tcaagtgcct 1260
ctctttcact atggcatata aagaggccag ggcttgaaa atgctcaggt gcatttcagt 1320
tttgagctta taaaactggg tagataacat gactagttag caaaaatggc tttcactggt 1380

```

<210> 113
 <211> 393
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (163)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (191)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (198)
 <223> a, c, g or t

<220>
 <221> unsure
 <222> (206)
 <223> a, c, g or t

```

<400> 113
gcactgcagc cttaaacagg cacaattttg tctatacttc cagaacctag attattctga 60
aatctttaac aacataaggt ttagatacca tttgcattga gtaccacta ggtgccgact 120
cttttaaggt gcatttttag tttcattatc tcaactttgt aangttggca tcattattcc 180
cattttacag nagataanat tgaagnaaag tcaagtttag gggattttca aggttgatca 240
gtacaactgg gtgacaaaat ttttgcctct tcaatgataa tgaggcctct gacatcttcc 300
tttctcataa gactacattt agtataactt atatatttta tcagtcaaca actatctttt 360
gagaacttgt acaccagga ctgtgtaatg ggc

```

<210> 114
 <211> 440
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (95) .. (291)
 <223> a, c, g or t

<400> 114
 gtccttttat tattcttttt ttcttatatt tttattgtgg cacaaacttt atactaaaag 60
 gaaaaggaga ttcatgtgta atacatgata agtgnnnnnn nnnnnnnnnn nnnnnnnnnn 120
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn natgatatca 300
 aactagggtc gtccctgccca cgtgcagcaa gccaatcact atgatgatgg gttttgccaa 360
 aagagacaag attttattca tagggctgct gaatgaggag acaggagagc aaatcccaa 420
 tctggcacc c tgaaaatagg 440

<210> 115
 <211> 791
 <212> DNA
 <213> Homo sapiens

<400> 115
 gaaatccaaa caactgccat tgatttattc atttatttca caaatattta ctgaacgcat 60
 ccagcatgct ctgtgggggt ctgtgctggg gctgggggtg ccaggatgag aaacagccgt 120
 gtggtgtgct tcttggttcc accagccaga cgagtgttgc ctttgcaagg agaaaggact 180
 cacaaggctt acacatttgc tgccctcagt tttgcccttt ctcaaataaa tctcacacat 240
 ccaatctcct tgttgcccat tagggagtat ataataaat taagtaaatg aggaattgcc 300
 taaaactaag ggagtttcac ctccatgtag gtagaagaat gtgaaatggg ctgtgtccag 360
 aagccagatc agaaatgggc catagcaagg tggggagggc agcgggtacc cacctggcag 420
 tgtagggggg ttgattcagc ttcattcttc tgacccttg tcaagtggac aagctccagc 480
 caaacaagg aagtgtgttg gaggggccac cagcacagaa gtgtaccttt ctgggtaatg 540
 tgtcaccagc tcccctggcc atgtgagagg acaggcacag ttgccacaca gtactaatag 600
 ttggtctctt ctttaagggt caaaaaaag gaggtggagc acttttaaga aagtgttaag 660
 gttccatgaa gatgttatgg tggcgtgctg gcaggtgcat atcaaccctg ccctgaggcc 720
 ctacgcagcc ttcggtctcc ccaaagcaat atggctcctt ataaagaagt cttttagggc 780
 tgggctcggt a 791

<210> 116
 <211> 4351
 <212> DNA
 <213> Homo sapiens

<400> 116
 cggcccagca aagttcttga caaagtaaag aagccaatta caagaaaatc tcagtttgca 60
 aagggacaga attcgtctca cttattcact gcttattgtg tgccctctct gttttttctt 120
 cttgtgggtc ggtgtatttg atgctggtta gtagagacaa agaagaagga caaacaggat 180
 aaagggtggat ctttggtgtg gaccctctgc actgcgaaag aagccacatc accgccaatg 240
 tggaaaatat gcaaagtgcc gtttaggaaga aggaaggata tgtgtgcagc atatgaagtg 300
 ccttgaatac gattaacttc cttcatgag tagtaaatag tagataactc tgatcaaaaa 360
 agggattcat gtgatttatc aagctgagca actgcgcgtc tgcagagaag ctggagggtc 420
 atcttgaat ctagggcaag aggagcacta ggcaattgcc aggactaaga agttaatcat 480
 acccttggtc tgcttccatc tgtctcagag tgacagcgtc gctctcagcg agcaggcatg 540
 ctttatagca gcagatcagg aattaatatt ttctgtgaaa cctcaagcat catttgcatg 600
 aacttggtt ttataaaaat ggaacataat tttatatgaa taaatcacgt tcagctagaa 660
 atacgagagg ctgcaaaaaa ttatgcttga cttaaaaaaa aagagagagg aacgagcaaa 720
 aaagccaaca tgaaaacagt tgttgaagcg atggcacttg gagggcacag atagccatgt 780
 ggttaaatg tgcatataat catctgaaat gtcagcctgc gactccagca acggctacat 840

ttttatagcc	ttggagatgc	atcaccagac	tgtaacctgt	ctgcccacat	ggcagtatct	900
caggcccaag	atccatttat	attgaaatcc	aaacaactgc	cattgattta	ttcatttatt	960
tcacaaatat	ttactgaacg	catccagcat	gctctgtggg	gtgctgtgct	ggggctgggg	1020
gtgccaggat	gagaaacagc	cgtgtggctg	tgctcttggc	ttcaccagcc	agacgagtg	1080
tgcccttgca	aggagaaagg	actcacaagg	cttacacatt	tgctgccctc	agttttgccc	1140
tttctcaaat	aaatctcaca	catccaatct	ccttggtgcc	cattagggag	tatataatga	1200
aattaagtaa	atgaggaatt	gcctaaaact	aaggaggttt	cacctccatg	taggtagaag	1260
aatgtgaaat	ggctctgtgc	cagaagccag	atcagaaatg	gtccatagca	aggtggggag	1320
ggcagcgggt	acccacctgg	cagtgtaggg	gggtggattc	agcttcatct	tcctgacccc	1380
ttgtcaagtg	gacaagctcc	agccaaacaa	aggaagtgtg	ttggagtggc	caccagcaca	1440
gaagtgtacc	tttctgggta	atgtgtcacc	cagtccctcg	gccatgtgag	aggacaggca	1500
cagttgccac	acagtactaa	tagttggctc	cttctttaag	ggtcaaaaaa	aaggaggtgg	1560
agcactttta	agaaagtgtt	aagggtccat	gaagatgtta	tggtggcgtg	ctggcagggtg	1620
catatcaacc	ctgccctgag	gccctcagca	gccttcggtc	tccccaaagc	aatatggctc	1680
cttataaaga	agtcttttag	ggctgggctc	ggtagctcat	gcctgtaatc	ccagcacttt	1740
gggaggccaa	gggaagaaat	ttgatctcaa	attcagtcac	ccagtctatt	caactgagga	1800
ccctcagca	ggggttccct	gtgctaaggc	cctgggtggc	gtagccagtg	gagagaaggc	1860
tctagccttg	atacctctcc	cttgggttaa	acgtaattcg	aagaagccta	agctctctgc	1920
gcagcccgcc	gcgcagcccg	ccgccccagc	ctgggcgaag	cccctgacgg	accaggagaa	1980
gcggcgccag	atcagcatcc	gcggcatcgt	gggcgtggag	aacgtggcag	agctgaagaa	2040
gagtttcaac	cggcacctgc	acttcacgct	gggtcaaggac	cgcaacgtgg	ccaccacccg	2100
cgactactac	ttcgcgctgg	cgcacacggt	gcgcgaccac	ctgggtggggc	gctggatocg	2160
cacgcagcag	cactactacg	acaagtgcc	caagagggtg	tattacctct	ctctggaatt	2220
ttacatgggc	cgaacattac	agaacacccat	gatcaacctc	ggtctgcaaa	atgcctgtga	2280
tgaggccatt	taccagcttg	gattggatat	agaagagtta	gaagaaattg	aagaagatgc	2340
tggacttggc	aatggtggtc	ttgggagact	tgctgcctgc	ttcttggtat	ccatggcaac	2400
cctgggactt	gcagcctatg	gatacggcat	tcggtatgaa	tatgggattt	tcaatcagaa	2460
gatccgagat	ggatggcagg	tagaagaagc	agatgattgg	ctcagatatg	gaaacccctg	2520
ggagaagtcc	cgcccagaat	tcattgctgcc	tgtgcacttc	tatggaaaag	tagaacacac	2580
caacaccggg	accaagtggg	ttgacactca	agtggctcctg	gctctgccat	atgacacccc	2640
cgtgcccggc	tacatgaata	acactgtcaa	caccatgcgc	ctctggtctg	ctcgggcacc	2700
aaatgacttt	aacctcagag	actttaatgt	tggagactac	attcaggctg	tgctggaccg	2760
aaacctggcc	gagaacatct	cccggtcct	ctatcccaat	gacaatgtgg	ccatccagct	2820
gaatgacact	cacctgcac	tcgcgatccc	tgagctgatg	aggatttttg	tggatattga	2880
aaaactgcc	tggtccaagg	catgggagct	caccagaag	acctcgccct	acaccaacca	2940
cacagtgtct	ccggaagccc	tggagcgctg	gcccgtggac	ctgggtggaga	agctgctccc	3000
tcgacatttg	gaaatcattt	atgagataaa	tcagaagcat	ttagatagaa	ttgtggcctt	3060
gtttcctaaa	gatgtggacc	gtctgagaag	gatgtctctg	atagaagagg	aagggaagcaa	3120
aaggatcaac	atggcccac	tctgcattgt	cggttcccat	gctgtgaatg	gcgtggctaa	3180
aatccactca	gacatcgtga	agactaaagt	attcaaggac	ttcagtgage	tagaacctga	3240
caagtttcag	aataaaacca	atgggatcac	tccaaggcgc	tggctcctac	tctgcaacce	3300
aggacttgca	gagctcatag	cagagaaaat	tggagaagac	tatgtgaaag	acctgagcca	3360
gctgacgaag	ctccacagct	tcctgggtga	tgatgtcttc	ctccgggaac	tcgccaagg	3420
gaagcaggag	aataagctga	agttttctca	gttctgggag	acggagtaca	aagtgaagat	3480
caaccatcc	tccatgtttg	atgtccaggt	gaagaggata	catgagtaca	agcgacagct	3540
cttgaactgt	ctgcatgtga	tcacgatgta	caaccgcatt	aagaaagacc	ctaagaagtt	3600
attcgtgcca	aggacagtta	tcattgggtg	taaagctgcc	ccaggatata	acatggccaa	3660
aatgatcata	aagctgatca	cttcagtggc	agatgtgggtg	aacaatgacc	ctatgggttg	3720
aagcaagtgt	aaagtcatct	tcttgagaa	ctacagagta	tctcttgctg	aaaaagtc	3780
tccagccaca	gatctgtcag	agcagatttc	cactgcaggc	accgaagcct	cggggacagg	3840
caatatgaag	ttcatgctaa	atggggccct	aactatcggg	accatggatg	gggccaatgt	3900
ggaaaatggca	gaagaagctg	gggaagagaa	cctgttcac	tttggcatga	ggaatagatga	3960
tgtggctgct	ttggacaaga	aagggtacga	ggcaaaaaga	tactatgagg	cacttccaga	4020
gctgaagctg	gtcattgatc	aaattgacaa	tggctttttt	tctcccaagc	agcctgacct	4080
cttcaagat	atcatcaaca	tgctatttta	tcatgacagg	tttaagtct	ttgcagacta	4140
cgaagcctat	gtcaagtgtc	aagataaagt	gagtcagctg	tacatgaatc	caaaggcctg	4200
gaacacaagt	gtactcaaaa	acatagctgc	ctcggggaaa	ttctccagtg	accgaacaat	4260
taaagaatat	gccccaaa	tctggaacgt	ggaaccttca	gatctaaaga	tttctctatc	4320
caatgaatct	aacaaagtca	atggaaattg	a			4351

<210> 117
 <211> 454
 <212> DNA
 <213> Homo sapiens

<400> 117
 tgtcaataca atcgggggaa aggaatactt tgaactactt tgttggaagg agtttgaaat 60
 cgttgaggac tcagcagcat gaagtagaga aattcacaat tggtagaaag gactattgtc 120
 cttcaacctt cattaagggtt aactattcaa cttcatttaa aaacagaaaag tgacaatttc 180
 acagcaaatt ctagaacttt agatcaaaaag tcaactcaat atggggggatt tatataagaa 240
 agagttaaaa aaaagacgaa atgtaatatc tatgttattg caagtgaag gaaaacagga 300
 agataaatat cacaagaaga caaaaatgta tctaactttt tgggacaaga ttgtggggtc 360
 cacagaaaat tgggaacttg aacttcctgt tccacagaga taaganatac cttgctttta 420
 tctcacttct caaaaaagta agtgatgggg ttag 454

<210> 118
 <211> 504
 <212> DNA
 <213> Homo sapiens

<400> 118
 tgtcaataca atcgggggaa aggaatactt tgaactactt tgttggaagg agtttgaaat 60
 cgttgaggac tcagcagcat gaagtagaga aattcacaat tggtagaaag gactattgtc 120
 cttcaacctt cattaagggtt aactattcaa cttcatttaa aaacagaaaag tgacaatttc 180
 acagcaaatt ctagaacttt agatcaaaaag tcaactcaat atggggggatt tatataagaa 240
 agagttaaaa aaaagacgaa atgtaatatc tatgttattg caagtgaag gaaaacagga 300
 agataaatat cacaagaaga caaaaatgta tctaactttt tgggacaaga ttgtggggtc 360
 cacagaaaat tgggaacttg aacttcctgt tccacagaga taagaaatac acttgctttt 420
 atctcacttc tcaaaaaaag taagatgaat ggggttttag gcccagaga cggacattgt 480
 agctgcaatc aattgtacta tctg 504

<210> 119
 <211> 407
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (385)
 <223> a, c, g or t

<400> 119
 aaaaaacagt ttggctatgt ttcagaagtc aaaaataagt ctgtaacctt tgacccagta 60
 atcctatttc tgggaagtcta aattgagaaa atgtggggta ctgaaaatct ctatttgcatt 120
 gaatatattt ataataacat tcgttatatt ctttatattc ataaaacatt ggaaacaatt 180
 tttatggcca aaaatggatg aatagctcag taaatgacgg ttctctgcaa gcgatgtaatt 240
 agtatgcagt cagtaagcaa atacagaaga tactaagttg caacattaga atatataata 300
 ttgtgtatta ggaagtcagg ttatcatatt taaattttga acaaaaagtaa aggttagatc 360
 agttcaattg agaaataggg gtcanttcag aaaatgttat tccatga 407

<210> 120
 <211> 104
 <212> DNA
 <213> Homo sapiens

<400> 120

taaaagaagtg ggtatcaggg actcctgtga gatagcatga gaaggtggta catttgggag 60
gtctcaaggg gttactgaat tattggaatt agaatcaaag ggac 104

<210> 121

<211> 149

<212> DNA

<213> Homo sapiens

<400> 121

tacagcaata gataattaat acttaattat ctaattaata catattaata ttttggcaac 60
atacactatg ttcctaaggt acctcggaat atcctcagaa ccatgtgttg caaatggcaa 120
tgctgtggta caatggggtc tcctaggca 149

<210> 122

<211> 419

<212> DNA

<213> Homo sapiens

<400> 122

ggaaatgtgt ttagttgtca tataaaagga aaatgcagtt taaaataatt tcagtaattg 60
cattcttgag ttttctgtcc tccctggtag catgaaactg gagatctttg gagacctatc 120
acagaacatg tactggaatt gtttgtgtgt ggagtaaagg cagctgtttg tagccatcta 180
gttgggaact gtctttcctt ggatagttag ctactctgtt ggtgtgtggg gtaacactta 240
cctgttgctg gcacgtagtc agtgatttct gtcattgtata agtaggcctt gccattgtca 300
gcaggtaatg atcttggaat gaccaacttc tgtaaatgta atccacaatc tagtgagggg 360
attatagcta tcaaacatat ttctcagtc actttttaag aagtagtcat ttaggctgg 419

<210> 123

<211> 691

<212> DNA

<213> Homo sapiens

<400> 123

aaagagacag ggtcttgctc tgtcaccag gctggagtag aatgacgtaa taatagetca 60
ccgcaacttc gaactcccgg gctcaagcaa tccttctgcc tcagcctccc aagagctggg 120
actacagaca tgtgctacca catccagctt ttttattttt ttagagagta gggctcctc 180
atggttgcca ggtgggtctc aactccacc tcaagcaatc ctacagcttc agcctcccaa 240
agagctagaa ttacaggcct gagccactgc acccagccta aatgactact tcttaaaaag 300
tggactgaga aatatgtttg atagctataa tcccctcact agattgtgga ttacattaac 360
agaagttggg ctttccaaga tcattacctg ctgacaatgg caaggcctac ttatacatga 420
cagaaatcac tgactacgtg ccagcaacag gtaagtgtta caccacacac caacagagta 480
gctaactatc caaggaaaga cagttcccaa ctatagggct acaaacagct gcctttactc 540
cacacacaaa caattccagt acatgttctg tgataggtct ccaaagatct ccagtttcat 600
ggtaccaggg aggacagaaa actcaagaat gcaattactg aaattatttt aaactgcatt 660
ttccttttat atgacaacta aacacatttc c 691

<210> 124

<211> 476

<212> DNA

<213> Homo sapiens

<400> 124

tagcacgtcg taaacgatga atagatatta gctttaaaaa tgatacttgt tattctgtgt 60

47

```

gctagatatc tagggaagtg aaggaaggac ggcaagggag gcagagatga ataaggcagt 120
gactaggccc catgggaggg agatcgcggt accacagctg aatggattgt cccccctaca 180
ttgccattca gctaagagac attcagcaat ttattgaata agcacttctt gagcccctag 240
tgcatgcata agacactgcg ttagggtctg gtgcacagca gtgaataaga cagacgtagt 300
tcttgctctc gagtgtctcat ggtccaatga gggagacaga ggggtgactgg gaacaacagt 360
ccagtgtgat aatgctagca tagcagcaga acaggggctg cacaacacaa aagaaggaa 420
atctaactcc caaatgaaaa gaggggcatt gacaaagtcc tcctagggaa aaagaa 476

```

<210> 125

<211> 491

<212> DNA

<213> Homo sapiens

<400> 125

```

cccttagaat aatgtctagc acgtcgtaaa cgatgaatag atattagctt taaaaatgat 60
acttgttatt ctgtgtgcta gatatactagg gaagtgaagg aaggacggca agggaggcag 120
agatgaataa ggcagtgact agggcccatg ggagggagat cgcggtacca cagctgaatg 180
gattgtctcc cctacattgc cattcagcta agagacattc agcaatttat tgaataagca 240
cttcttgagc ccctagtgc tgcatacagac actgcgttag ggctgggtgc acagcagtga 300
ataagacaga cgtagttctt gctctcgagt gctcatggtc caatgagggg gacagagggg 360
gactgggaac aacagtccag tgtgataatg ctagcatagc agcagaacag gggctgcaca 420
aacacaaaga aggaacatct aactcccaaa tgaaaagagg ggcattgaca aagtcctcct 480
agggaaaaag a 491

```

<210> 126

<211> 752

<212> DNA

<213> Homo sapiens

<400> 126

```

ctcagctgag aagcagacac attgtgaaat ggactcccc aaaagagttt catctgactt 60
atcccttctc cgcaataaaa tcttggaatc tgggtgtggt tggttttagat gctgtgggtac 120
cggctgggtt tagcaacaag gacagtgttg gtagggtgag aaacactatc ccaagtcata 180
tgtctgtgtg actacaggac atttcttttg aatgccacaa ggatgattta tatgattact 240
ggtgacaagc ctctgtctcc tgaagacagg ccaagataac gttagattga atttcaagag 300
atgaaagtga ggtttttaag taatagcaaa gccttgtggt tctgtagtac tttgtgcttt 360
ttgaagtgtc ttcacagtca ttatcctgtt tgatcctact aagaaccctg aaagtacata 420
ggttgtgtgt ttttatcctg agactacaaa tgataccaag gataacgatg agtaggaatc 480
agagctagaa ttaaccctta ttttcttact attgaccag catgctttct atgttgaaaa 540
gtgcaccaca tcgagaagag attggtcacc gcagcacagg gcacgcagaa ttccattagt 600
atcacttacc tgggaagtcc aggtgccttc aatagttgag gggagtaaat gatatgacta 660
cctaccttca aaacttgtag tttaaagtgg taacttgaat actcacattt acctctgttt 720
ccttcctcta aaagaatggg tttttaagg gt 752

```

<210> 127

<211> 158

<212> DNA

<213> Homo sapiens

<400> 127

```

aaaaaaaaa aaaaagacag ttgggtgtgc atatctcttc tgccctttaat ttgttgaggt 60
acctcatgtg tagccttttg aatactcttc tgtatactgg tgagagaatt agagtgaata 120
aagcagataa catcttagtg ttattaatga aagtagta 158

```

<210> 128

<211> 642
 <212> DNA
 <213> Homo sapiens

<400> 128
 tttatttggg tttccagctt tactgcaggt atgattgaca aataatgtct gtttgtaaaa 60
 tttcagtcga gtcataagata ccaggttaagg cagagagtgg gagggagact gaggccttgg 120
 tctggtggtg ggagcactgc agctcgagtc ttggagtcag gaggggggtg ttgcacttcc 180
 ctgtctctgct cctttttcag ctttctggtt ccctgtagct tctggaactg attattttttg 240
 tttctttaat gctgccctgt cttgtaaaag gagagccatt agcatcattt gttttcagga 300
 gagaagcaga tttgaaggct caggaacttc ctgggaaagg tgacctcttt tgagccaaga 360
 gctttacccc ctagtttttt gttttttttt tctctgtct acctggagct gagaggttat 420
 ccctttcaat ccctctcaag gtccagaatc accagctagg gttgggtctg ccctggagc 480
 acagactcct cccttgggga cccagagcc cttatcagta tatcagtaag agggcaagag 540
 aacagagatt gtcagagcag aggaaacgtg tattctgtgc ccagcccca ctccatgaat 600
 attccctgtg ctcaaagcac atacttaggc taagaacagg at 642

<210> 129
 <211> 220
 <212> DNA
 <213> Homo sapiens

<400> 129
 cttttcttgg ggagaatttt tttttttatt tttagcttcc gattcttata gaaatgtaat 60
 actaggcgat tcataattat atagacaagt ttttctgaaa tggttcatttg ttcattttatc 120
 atttttaacc cagtctgctt ctaacaggtc ataagttaca ttccaagata tggatatgat 180
 aaaactattg aatgaagtat taaaagaatc aagttcatgg 220

<210> 130
 <211> 507
 <212> DNA
 <213> Homo sapiens

<400> 130
 tcattttgta tgaaagggga attttaggaa ttagctggag atagacattt gggaatagct 60
 aggataaaga tagtaattgc tgattcacca aaacaaaaag aagtgttaga tttgaaaatt 120
 ttgtaggaaa ccaccaggtt ctcacctctt gtggtgtgtg tgtatgtgct gtattttttt 180
 ttaaaactact gaaaactcaa gatctttgtt gttccacaga ttcagttctg tgtcttgtct 240
 aattatgccc caggtatatg ataatgtaca gtcacgtttc ttagagtaac tcagaacatt 300
 tatgacacag gggttatctt acttctctag tctcagagtt tcacttagca ggtcatctga 360
 gtgaaatcta agccagattc ctgtggatct taatgaaaag gtagtagaaa gtagtggcat 420
 agcttgaaat ttaactattg tcagatattg gggcaaaaac catctgtata cctcatgggc 480
 ctccagtaaa cacttgtaaa ttatgag 507

<210> 131
 <211> 760
 <212> DNA
 <213> Homo sapiens

<400> 131
 tcattttgta tgaaagggga attttaggaa ttagctggag atagacattt gggaatagct 60
 aggataaaga tagtaattgc tgattcacca aaacaaaaag aagtgttaga tttgaaaatt 120
 ttgtaggaaa ccaccaggtt ctcacctctt gtggtgtgtg tgtatgtgct gtattttttt 180
 ttaaaactact gaaaactcaa gatctttgtt gttccacaga ttcagttctg tgtcttgtct 240
 aattatgccc caggtatatg ataatgtaca gtcacgtttc ttagagtaac tcagaacatt 300
 tatgacacag gggttatctt acttctctag tctcagagtt tcacttagca ggtcatctga 360

49

```

gtgaaatcta agccagattc ctgtggatct taatgaaaag gtagtagaaa gtagtggcat 420
agcttgaaat ttaactattg tcagatattg gggcaaaaac catctgtata cctcatggac 480
ctccagtaaa cacttgtaga ttatgagttt agattgttta aagtagattt cagtatttcc 540
agagtgaatt tagtgttact tgtgaggagg aggggtgagaa tatgtatcta gttgagtggg 600
agtacttgtg tgtctacggg tcgtaacggc catgcaacac caccacgga atcgagaaag 660
agtataaatc tgtcaatcct gtacgtgtcc ggaccgagtg aggtttcccg tgttgagtaa 720
aattaagccg cattctccac tcctggtgtt gcctaacgtc 760

```

<210> 132
 <211> 214
 <212> DNA
 <213> Homo sapiens

```

<400> 132
caagatttgg ggcaaggaga ccagtttaga ggactaatcc agaagatgga tattgatgat 60
ttctacttag agatttagaa agaagactcg agtacctagc ttttcatgtc tctgtatttg 120
ttttctcctt ttcactgccc ttttttcttc cctcatttac ccctgtgttc tgtactgtca 180
cttgcttcca gttgtcaata tgttgatttc tgtt 214

```

<210> 133
 <211> 479
 <212> DNA
 <213> Homo sapiens

```

<400> 133
ccttaggata aaaattagtc ttcccaacag gagatacaaa gaccaccaga actggttcag 60
ttcctggctc tccattcaca tcattcattt tctctacctc agacttgaca ctccagtata 120
actttttgtt gatagtagtt cagtgggata gaccatcaat tgattgcata cctccatgct 180
ttgctaattg tcttctatct atccaaaacc cttcccatgt ttttgcttaa acatcattca 240
tattccaaga cttaaagtcaa tgaaaatcta tatcaggatg attgtectca atcttctggg 300
tggtactacat gtctctcatc aattatactt tgtatcatca gtctgattca ttcaaatagt 360
ctgtgtatta tatgtgcctc aggctaata ctattaatac ctgtatatta gaaaagaaag 420
cctggtgctt agtagaattt tgttaaatat ttgctcagct gaaccaatgc attaatact 479

```

<210> 134
 <211> 270
 <212> DNA
 <213> Homo sapiens

```

<400> 134
tagggatttc gtcacttgga agtaagaagg ttcagtcatc tttggccagc tttgtgttgt 60
gttgaaaatt agcccccata gagaattcct gcagaaggtc agggctcttg gggatattct 120
acacttgagc ctctttcttt ttttaagatga catacttggt atagttgtca aatatggaca 180
ataacaggaa gccaaactca aataataata atagggtgtt acaaagccgt ggcacatggt 240
cccactgta gtccagctgt ctggagctga 270

```

<210> 135
 <211> 404
 <212> DNA
 <213> Homo sapiens

```

<400> 135
acgcgtccgt gaaaaggaag aataacctatt acttaggtat tgggaaattg aaaatgaaga 60
atggaagaaa gagggaggga agagactgtt gtgtttctat ggagaacaac attggggccc 120
ttgactttag atttcagtgg ggacctacaa aaaggaaaaa tggaaaggga attctgaagt 180

```

50

cttaaggtgg gctatctgaa agttggatcc ctgggtgaaa aagattttat aatattagat 240
 gagttgagag aaccaatgtg aattaaagct gactggctta aaaaaataa acccatcaaa 300
 attagtaagg gaataatgtt attcattgcc ttttttctg tgagttatga aagctcttcg 360
 aagatgaagg ttttatgaaa ctcaagatct ctccagaggc cggg 404

<210> 136
 <211> 553
 <212> DNA
 <213> Homo sapiens

<220>
 <221> unsure
 <222> (446)
 <223> a, c, g or t

<400> 136
 acgcgtccgt gaaaaggaag aatacctatt acttaggtat tgggaaattg aaaatgaaga 60
 atggaagaaa gagggagggg agagactgtt gtgtttctat ggagaacaac attggggccc 120
 ttgacttttag atttcagtgg ggacctacaa aaaggaaaaa tggaaagggg attctgaagt 180
 cttaaggtgg gctatctgaa agttggatcc ctgggtgaaa aagattttat aatattagat 240
 gagttgagag aaccaatgtg aattaaagct gactggctta aaaaaataa acccatcaaa 300
 attagtaagg gaataatgtt attcattgcc ttttttctg tgagttatga aagctcttcg 360
 aagatgaagg ttttatgaaa ctcaagatct ctccagaggc cgggcacagt ggctcgcgcc 420
 tgtaattcca gcactttggg aggctnagg gagcagattg cgagtccaga agtgagcaga 480
 ttgcttgagt ccaggagttc gagaccagcc tgggcaacat ggcaaaaccc ctgtctctac 540
 taaaaaaaaa aaa 553

<210> 137
 <211> 41
 <212> PRT
 <213> Homo sapiens

<400> 137
 Met Lys Val Arg Ser Ile His Pro Ser Ser Ala Thr Cys Ala Ser Ala
 1 5 10 15
 Leu His Leu Pro Gln Leu Thr Thr Glu Lys Arg Thr Gln Leu His Lys
 20 25 30
 Arg Asp Cys Lys Ile Arg Lys Tyr Ile
 35 40

<210> 138
 <211> 47
 <212> PRT
 <213> Homo sapiens

<400> 138
 Met Val Thr Leu Gln Met Pro Ser Val Ala Ala Gln Thr Ser Leu Thr
 1 5 10 15
 Asn Ser Ala Phe Gln Ala Glu Ser Lys Val Ala Ile Val Ser Gln Pro
 20 25 30
 Val Ala Arg Ser Ser Val Ser Ala Asp Ser Arg Ile Cys Thr Glu
 35 40 45

<210> 139
 <211> 55
 <212> PRT
 <213> Homo sapiens

<400> 139
 Ile Gln Asp Lys Asp Ser Val Asn Met Val Thr Leu Gln Met Pro Ser
 1 5 10 15
 Val Ala Ala Gln Thr Ser Leu Thr Asn Ser Ala Phe Gln Ala Glu Ser
 20 25 30
 Lys Val Ala Ile Val Ser Gln Pro Val Ala Arg Ser Ser Val Ser Ala
 35 40 45
 Asp Ser Arg Ile Cys Thr Glu
 50 55

<210> 140
 <211> 47
 <212> PRT
 <213> Homo sapiens

<400> 140
 Met Phe Leu Tyr Ala Phe Met Tyr Ile Phe His Leu Tyr Asn Glu Cys
 1 5 10 15
 Met Tyr Leu Leu Ser Leu Tyr Lys Leu Leu Leu Phe Val Ile Phe Phe
 20 25 30
 Phe Phe Pro Phe Phe Gly Phe Leu Thr Phe Gln Lys Met Lys His
 35 40 45

<210> 141
 <211> 70
 <212> PRT
 <213> Homo sapiens

<400> 141
 Met Asn Leu Gly Asn Lys Pro Tyr Phe Leu Ile Thr Met Leu Asp His
 1 5 10 15
 Leu Ser Pro Arg Arg Gly Trp Gly Thr Gln Asp Glu Ser Leu Gly Ser
 20 25 30
 Leu Trp Tyr Gln Ile Leu Asn Ile Pro Ser Leu Leu Asn Ala Thr Leu
 35 40 45
 Leu Leu Pro Leu Leu Glu Gly Lys Asn Ala Lys Met Gly Ile Ser Leu
 50 55 60
 Ser Leu Gly Pro Val Pro
 65 70

52

<210> 142
 <211> 11
 <212> PRT
 <213> Homo sapiens

<400> 142
 Met Tyr Trp Tyr Ser Phe Gln Ser Ser Ser Trp
 1 5 10

<210> 143
 <211> 230
 <212> PRT
 <213> Homo sapiens

<400> 143
 Leu Asp Arg Leu Ser Lys Ala Lys Ile Asp Lys Lys Thr Leu Asp Leu
 1 5 10 15
 Asn Ala Thr Leu Asp Gln Met Asp Leu Thr Asp Ile Tyr Arg Thr Val
 20 25 30
 Tyr Leu Thr Pro Thr Asp Tyr Thr Phe Phe Ser Ser Ala Cys Gly Thr
 35 40 45
 Phe Ser Arg Ile Asp His Met Leu Ser His Lys Thr Ser Leu Asn Lys
 50 55 60
 Phe Leu Lys Ile Gly Ile Ile Gln Ser Ile Phe Ser Asp His Lys Arg
 65 70 75 80
 Ile Lys Leu Glu Ile His Thr Lys Arg Asn Phe Gly Asn Tyr Thr Asn
 85 90 95
 Thr Trp Lys Leu Asn Met Leu Leu Asn Asn Tyr Trp Val Asn Glu Glu
 100 105 110
 Ile Lys Met Glu Ile Ala Lys Phe Leu Lys Thr Asn Arg Asn Gly Asn
 115 120 125
 Ala Thr Tyr Gln Asn Met Trp Asp Thr Ala Arg Ala Met Ala Arg Gly
 130 135 140
 Asn Leu Thr Val Ile Asn Ala Tyr Ile Lys Lys Val Val Glu Ile Phe
 145 150 155 160
 Ala Ile Lys Asn Leu Ser Met His Leu Lys Glu Leu Glu Lys Gln Lys
 165 170 175
 Gln Thr Asn Pro Gln Ser Ser Arg Gln Lys Glu Ile Met Lys Ser Arg
 180 185 190
 Ala Asp Gln Asn Glu Thr Asp Lys Lys Thr Ile Gln Arg Val Asn Glu
 195 200 205
 Met Lys Ser Cys Phe Phe Lys Lys Ile Asn Lys Ile Asp Asn Pro Leu
 210 215 220
 Ala Ala Leu Thr Lys Lys

53

225

230

<210> 144
 <211> 149
 <212> PRT
 <213> Homo sapiens

<400> 144
 Met Tyr Gln Leu Arg Leu Val Thr Leu Phe Gln Ile His Met Lys Gly
 1 5 10 15
 Ala Ile Pro Leu Lys Leu Phe Thr Asp Val Leu Cys Lys Arg Trp Ser
 20 25 30
 Thr Lys Glu Thr His Gln Met Gly Gly Glu Ala Asp Pro Gly His Ala
 35 40 45
 Gln Arg Glu Gln Leu Gly Thr Trp Ala Gly Ile Gly Lys Lys Val Val
 50 55 60
 Gln Arg Ala Arg Pro Gly Pro Ala Leu Ser Gly Gly Ser Gly Gly Leu
 65 70 75 80
 Cys Leu Ser Ala Leu Pro Pro Gly Leu Pro Pro Met Thr Val His Pro
 85 90 95
 Cys Arg Asn His Leu Arg Pro Pro Thr Pro Thr Pro Ala Pro Leu Gly
 100 105 110
 Ser Tyr His Leu Pro Phe Pro Pro Ser Ser Leu Ser Pro Thr Lys Ala
 115 120 125
 Ser Leu Cys Phe Leu Glu Ala Ser Ile Thr Gly Ser Cys Pro Gly Pro
 130 135 140
 Ser Trp Gly Thr Arg
 145

<210> 145
 <211> 31
 <212> PRT
 <213> Homo sapiens

<400> 145
 Met Gly Trp Asn Glu Glu Glu Gln Ser Cys Pro Pro Val Pro Gly Gly
 1 5 10 15
 Thr Val Ser Arg Lys Ile His Thr Tyr Leu Lys Leu Gln Lys Gly
 20 25 30

<210> 146
 <211> 106
 <212> PRT
 <213> Homo sapiens

<400> 146

54

Cys Gly Trp Trp Thr Gly Met Pro Gly Ser Ser Pro Gly Ser Leu Leu
 1 5 10 15
 Pro Ser Asn Arg Leu Ser Leu Val Pro Leu Val Pro Ser Ala Ser Met
 20 25 30
 Thr Arg Leu Met Arg Ser Arg Thr Ala Ser Gly Ser Ser Val Thr Ser
 35 40 45
 Leu Asp Gly Thr Arg Ser Arg Ser His Thr Ser Glu Gly Thr Arg Ser
 50 55 60
 Arg Ser His Thr Ser Glu Gly Thr Arg Ser Arg Ser His Thr Ser Glu
 65 70 75 80
 Gly Ala His Leu Asp Ile Thr Pro Asn Ser Gly Ala Ala Gly Asn Ser
 85 90 95
 Ala Gly Pro Lys Ser Met Glu Val Ser Cys
 100 105

<210> 147
 <211> 72
 <212> PRT
 <213> Homo sapiens

<400> 147
 Met Ser His Gly Ser Gly Trp Gln Cys Tyr Ser Pro Met Asn Thr Asp
 1 5 10 15
 His Ser Ser Asn Thr Gly Asp Trp Ser His Thr Ala Thr Phe Leu Ser
 20 25 30
 Arg Gln Arg His Lys Thr Arg Lys Asn Arg Thr Thr Leu Arg Ala Val
 35 40 45
 Met Trp Glu Cys Gly Pro Ser Tyr Asn Thr Gln His Gln Asn Trp Thr
 50 55 60
 Leu His Leu Lys Gly Phe Lys Thr
 65 70

<210> 148
 <211> 24
 <212> PRT
 <213> Homo sapiens

<400> 148
 Met Glu Gly Pro Thr Asn Arg Ser Ser Leu Glu Pro Pro Glu Glu Ala
 1 5 10 15
 Gln Pro Ser Gln Gln Phe Gly Arg
 20

<210> 149
 <211> 70

55

<212> PRT

<213> Homo sapiens

<400> 149

Met Leu Asp Leu Leu Ile Val Phe Arg Ile Lys Ser Lys Leu Leu Lys
 1 5 10 15

Met Ala Phe His Asp Leu Val Ser Pro His Gln Asn Ala His Thr Met
 20 25 30

Leu Leu Leu Thr Pro Ser Gln Leu Trp Leu Pro Ser Thr Cys Ser Ser
 35 40 45

Gln Ala Ser Thr Ser Phe Leu Val Ser Ala Val Leu Leu Ser Pro Pro
 50 55 60

Ser Leu Leu Ser Pro Gly
 65 70

<210> 150

<211> 46

<212> PRT

<213> Homo sapiens

<400> 150

Met Ser Thr Cys Phe Leu Ala Ser His Gly Asn Ser Cys Leu Leu Cys
 1 5 10 15

Ser Phe Ser Ile Ile Ser Leu Leu Leu Ala Ser Lys Glu Ser Phe Val
 20 25 30

Gly Ile Leu Pro Ser Ser Ser Tyr Leu Leu Cys Lys Ile Thr
 35 40 45

<210> 151

<211> 40

<212> PRT

<213> Homo sapiens

<400> 151

Met Glu Arg Phe Lys Glu Arg Gly Arg Gly His Gly Ala Phe Met Pro
 1 5 10 15

Ser Pro Gly Thr Leu Pro Ser Arg Asn Leu Gln Thr Val Gln Leu Ser
 20 25 30

Gly Ser Ser Leu Asn Leu Val Ile
 35 40

<210> 152

<211> 32

<212> PRT

<213> Homo sapiens

<400> 152

Met Leu Gly Ser Glu Cys Leu Leu Phe Met His Leu Leu Lys Lys Leu

56

1 5 10 15
 Leu Gln Gly Asn Lys Lys Arg Ile Gln Glu Arg Gly His His Gly Leu
 20 25 30

<210> 153
 <211> 956
 <212> PRT
 <213> Homo sapiens

<400> 153
 Met Lys Ala Glu Ile Lys Val Phe Phe Glu Thr Asn Glu Asn Lys Asp
 1 5 10 15
 Thr Thr Tyr Gln Asn Leu Trp Asp Thr Phe Lys Ala Val Cys Arg Gly
 20 25 30
 Lys Phe Ile Ala Leu Asn Ala His Lys Arg Lys Gln Glu Arg Ser Lys
 35 40 45
 Ile Asp Thr Leu Thr Ser Gln Leu Lys Glu Leu Glu Lys Gln Glu Gln
 50 55 60
 Thr His Ser Lys Ala Ser Arg Arg Gln Glu Ile Thr Lys Ile Arg Ala
 65 70 75 80
 Glu Leu Lys Glu Ile Gln Thr Gln Lys Thr Leu Gln Lys Ile Asn Glu
 85 90 95
 Ser Arg Ser Trp Phe Phe Glu Arg Ile Asn Lys Ile Asp Arg Ser Leu
 100 105 110
 Ala Arg Leu Ile Lys Lys Lys Arg Glu Lys Asn Gln Ile Asp Thr Ile
 115 120 125
 Lys Asn Asp Lys Gly Asp Ile Thr Thr Asp Pro Thr Glu Ile Gln Thr
 130 135 140
 Thr Ile Arg Glu Tyr Tyr Lys His Leu Tyr Ala Asn Lys Leu Glu Asn
 145 150 155 160
 Leu Glu Glu Met Asp Lys Phe Leu Asp Thr Tyr Thr Leu Pro Arg Leu
 165 170 175
 Asn Gln Glu Glu Val Glu Ser Leu Asn Arg Pro Ile Thr Gly Ala Glu
 180 185 190
 Ile Val Ala Ile Ile Asn Ser Leu Pro Thr Lys Lys Ser Pro Gly Pro
 195 200 205
 Asp Gly Phe Thr Ala Glu Phe Tyr Gln Ser Trp Ala Glu Thr Gln Pro
 210 215 220
 Lys Lys Glu Asn Phe Arg Pro Ile Ser Leu Met Asn Ile Asp Ala Lys
 225 230 235 240

Ile Leu Asn Lys Ile Leu Ala Lys Arg Ile Gln Gln His Ile Lys Lys
 245 250 255
 Leu Ile His His Asp Gln Val Gly Phe Ile Pro Gly Met Gln Gly Trp
 260 265 270
 Phe Asn Ile Arg Lys Ser Ile Asn Val Thr Gln His Ile Asn Arg Ala
 275 280 285
 Lys Asp Lys Asn His Met Ile Ile Ser Ile Asp Ala Glu Lys Ala Phe
 290 295 300
 Asp Lys Ile Gln Gln Pro Phe Met Leu Lys Thr Leu Asn Lys Leu Gly
 305 310 315 320
 Ile Asp Gly Thr Tyr Phe Lys Ile Ile Arg Ala Ile Tyr Asp Asn Pro
 325 330 335
 Thr Ala Asn Ile Ile Leu Asn Gly Gln Lys Leu Glu Ala Phe Pro Leu
 340 345 350
 Lys Thr Gly Thr Arg Gln Gly Cys Pro Leu Ser Pro Leu Leu Phe Asn
 355 360 365
 Ile Val Leu Glu Val Leu Ala Arg Ala Ile Arg Gln Glu Lys Glu Ile
 370 375 380
 Lys Gly Ile Gln Leu Gly Lys Glu Glu Val Lys Leu Ser Leu Phe Ala
 385 390 395 400
 Asp Asn Met Ile Val Tyr Leu Glu Asn Pro Ile Val Ser Ala Gln Asn
 405 410 415
 Leu Leu Lys Leu Ile Ser Asn Phe Ser Lys Val Ser Gly Tyr Lys Ile
 420 425 430
 Asn Val Gln Lys Ser Gln Ala Phe Leu Tyr Thr Asn Asn Arg Gln Thr
 435 440 445
 Glu Ser Gln Ile Met Ser Gln Leu Pro Phe Thr Ile Ala Ser Lys Arg
 450 455 460
 Ile Lys Tyr Leu Gly Ile Gln Leu Thr Arg Asp Val Lys Asp Leu Phe
 465 470 475 480
 Lys Glu Asn Tyr Lys Pro Leu Leu Lys Glu Ile Lys Glu Asp Thr Asn
 485 490 495
 Lys Trp Lys Asn Ile Pro Cys Ser Gly Glu Gly Arg Ile Asn Ile Val
 500 505 510
 Lys Met Ala Ile Leu Pro Lys Glu Leu Glu Lys Thr Thr Leu Lys Phe
 515 520 525
 Ile Trp Asn Gln Lys Arg Ala His Ile Ala Lys Ser Ile Leu Asn Gln
 530 535 540
 Lys Asn Lys Ala Gly Gly Ile Thr Leu Pro Asp Phe Lys Leu Tyr Tyr

58

545		550		555		560
Lys Ala Thr Val Thr Lys Thr Ala Trp Tyr Trp Tyr Gln Asn Arg Asp						
	565			570		575
Ile Asp Gln Trp Asn Arg Thr Glu Pro Ser Glu Ile Thr Gln His Ile						
	580		585			590
Tyr Ser Tyr Leu Ile Phe Asp Lys Pro Glu Lys Asn Lys Gln Trp Gly						
	595		600			605
Lys Asp Ser Leu Phe Asn Lys Trp Cys Trp Glu Asn Trp Leu Ala Ile						
	610		615			620
Cys Arg Lys Leu Lys Leu Asp Pro Phe Leu Thr Pro Tyr Thr Lys Met						
	625		630		635	640
Asn Ser Arg Trp Ile Lys Asp Leu Asn Val Arg Pro Lys Thr Ile Lys						
		645		650		655
Thr Leu Glu Glu Asn Leu Gly Ile Thr Ile Gln Asp Ile Gly Met Gly						
		660		665		670
Lys Asp Phe Met Ser Lys Thr Pro Lys Ala Met Ala Thr Lys Asp Lys						
	675		680			685
Ile Asp Lys Trp Asp Leu Val Lys Leu Lys Ser Phe Cys Thr Ala Lys						
	690		695			700
Glu Thr Thr Ile Arg Val Asn Arg Gln Pro Thr Lys Trp Glu Lys Ile						
	705		710		715	720
Phe Ala Thr Tyr Ser Ser Asp Lys Gly Leu Ile Ser Arg Ile Tyr Asn						
		725		730		735
Glu Leu Lys Gln Ile Tyr Lys Lys Lys Thr Asn Asn Pro Ile Lys Lys						
		740		745		750
Trp Ala Lys Asp Met Asn Arg His Phe Ser Lys Glu Asp Ile Tyr Ala						
	755		760			765
Ala Lys Lys His Met Lys Lys Cys Ser Ser Ser Leu Ala Ile Arg Glu						
	770		775			780
Met Gln Ile Lys Thr Thr Met Arg Tyr His Leu Thr Pro Val Arg Met						
	785		790		795	800
Ala Ile Ile Lys Lys Ser Gly Asn Asn Arg Cys Trp Arg Gly Cys Gly						
		805		810		815
Glu Thr Gly Thr Leu Leu His Cys Trp Trp Asp Cys Lys Leu Ala Gln						
		820		825		830
Pro Leu Trp Lys Ser Val Trp Arg Phe Leu Arg Asp Leu Glu Leu Glu						
	835		840			845
Ile Pro Phe Asp Pro Ala Ile Pro Leu Leu Gly Ile Tyr Pro Lys Asp						
	850		855			860

59

Tyr Lys Ser Cys Cys Tyr Lys Asp Thr Cys Thr Arg Met Phe Ile Ala
 865 870 875 880

Ala Leu Phe Thr Ile Ala Lys Thr Trp Asn Gln Pro Lys Cys Pro Thr
 885 890 895

Ile Ile Asp Trp Ile Lys Lys Met Trp His Ile Tyr Thr Met Glu Tyr
 900 905 910

Tyr Ala Ala Ile Lys Asn Asp Glu Phe Val Ser Phe Val Gly Thr Trp
 915 920 925

Met Lys Leu Glu Ile Ile Ile Leu Ser Lys Leu Ser Gln Glu Gln Lys
 930 935 940

Thr Thr His Arg Ile Phe Ser Leu Ile Gly Gly Asn
 945 950 955

<210> 154

<211> 39

<212> PRT

<213> Homo sapiens

<400> 154

Met Ile Ile Thr Ser Gln Gly Asn Phe Leu Phe Pro Leu Phe Ile Ser
 1 5 10 15

Leu Leu His His Tyr Ser Gln Ser Leu Ser Leu Phe Pro Lys Glu Val
 20 25 30

Phe His Gly Phe Leu Thr Asp
 35

<210> 155

<211> 37

<212> PRT

<213> Homo sapiens

<400> 155

Met Val Leu Ser Cys Tyr Ser Leu Val Thr Phe Arg Ser Ser Leu Leu
 1 5 10 15

Thr Lys Gly Lys Ile Ile Tyr Lys Tyr Gln Met Thr Ile Glu Leu Ser
 20 25 30

Gln Leu Met Phe Phe
 35

<210> 156

<211> 110

<212> PRT

<213> Homo sapiens

<400> 156

Met Gly Cys His Gly Gly Ala Arg Asp Ser Cys Val Asn Arg Glu Cys
 1 5 10 15

60

Gly Phe Leu Gln Arg Gly Val Trp Arg Trp Thr Ser Arg Ser Phe Trp
 20 25 30
 Ser Leu Arg Glu Gly Gln Gln Ser Ser Arg His Phe Met Asn His Ile
 35 40 45
 Leu Ala Val Ala Ala Phe Ala Ser Pro Gly Gly Trp Ser His Ala Leu
 50 55 60
 Ala Ala Arg Leu Arg His Pro Pro Val His Ser Val Pro Trp Pro Pro
 65 70 75 80
 Ala Val Gly Leu Ala Leu Phe Ser Thr Asn Asn Pro Gln Cys Ile Val
 85 90 95
 Met Thr Ser Ala Thr Asn Val Asp Val Ser Met Tyr His Ile
 100 105 110

<210> 157
 <211> 62
 <212> PRT
 <213> Homo sapiens

<400> 157
 Met Gly Ser His Phe Pro Gln Ser Arg Trp His Lys Leu His Glu Val
 1 5 10 15
 Ala Ala Val Pro Leu His Pro Asp Gln Ser Leu Ala Pro Gln Trp Asn
 20 25 30
 His Thr Pro Pro Leu Pro Glu Ala Glu Ser Leu Phe Tyr Gly Arg Ala
 35 40 45
 Ala Ala Leu Gly Thr Phe Leu Asn Ser Pro Val Phe His Leu
 50 55 60

<210> 158
 <211> 241
 <212> PRT
 <213> Homo sapiens

<400> 158
 Glu Gly Cys Leu Trp Pro Ser Glu Ser Thr Val Ser Gly Asn Gly Ile
 1 5 10 15
 Pro Glu Cys Pro Cys Cys Trp Asp Pro Pro Cys Arg Arg Ser Ser Ala
 20 25 30
 Pro Cys Pro Ala Gly Ser Ser Pro Ala Leu Cys Ser Leu His Thr Gly
 35 40 45
 Ala Arg Thr Leu Pro Leu Phe Gly Gly Gly Arg Pro Gln Val Tyr Ala
 50 55 60
 Pro Pro Arg Pro Thr Asp Arg Leu Ala Val Pro Pro Phe Ala Gln Arg
 65 70 75 80

61

Glu Arg Phe His Arg Phe Gln Pro Thr Tyr Pro Tyr Leu Gln His Glu
 85 90 95
 Ile Asp Leu Pro Pro Thr Ile Ser Leu Ser Asp Gly Glu Glu Pro Pro
 100 105 110
 Pro Tyr Gln Gly Pro Cys Thr Leu Gln Leu Arg Asp Pro Glu Gln Gln
 115 120 125
 Leu Glu Leu Asn Arg Glu Ser Val Arg Ala Pro Pro Asn Arg Thr Ile
 130 135 140
 Phe Asp Ser Asp Leu Met Asp Ser Ala Arg Leu Gly Gly Pro Cys Pro
 145 150 155 160
 Pro Ser Ser Asn Ser Gly Ile Ser Ala Thr Cys Tyr Gly Ser Gly Gly
 165 170 175
 Arg Met Glu Gly Pro Pro Pro Thr Tyr Ser Glu Val Ile Gly His Tyr
 180 185 190
 Pro Gly Ser Ser Phe Gln His Gln Gln Ser Ser Gly Pro Pro Ser Leu
 195 200 205
 Leu Glu Gly Thr Arg Leu His His Thr His Ile Ala Pro Leu Glu Ser
 210 215 220
 Ala Ala Ile Trp Ser Lys Glu Lys Asp Lys Gln Lys Gly His Pro Leu
 225 230 235 240
 Leu

<210> 159
 <211> 50
 <212> PRT
 <213> Homo sapiens

<400> 159
 Met Ile His Phe Leu Ser Phe Ser Thr Asn Asn Ala Tyr Ala Leu Asp
 1 5 10 15
 Leu Pro Glu Tyr Ser Trp Thr Thr Asp Leu Cys Lys Lys Leu Phe Phe
 20 25 30
 Leu Lys Ile Ala Ser Lys Gln Asn Gly Phe Asn Lys Leu Gln Asn Arg
 35 40 45
 Gln Pro
 50

<210> 160
 <211> 37
 <212> PRT
 <213> Homo sapiens

62

<400> 160

Met Ile Cys Pro Phe Phe Leu His Ser Phe Thr Ser Ser Ser Phe Tyr
 1 5 10 15
 Cys Tyr Phe Leu Lys Arg Ile Asn Pro Leu Ala Val Leu Phe Arg Val
 20 25 30
 Phe Phe Thr Leu Phe
 35

<210> 161

<211> 75

<212> PRT

<213> Homo sapiens

<400> 161

Met Leu Val Lys Ser Arg Cys Leu Cys Leu Cys Pro Phe Cys Leu Gly
 1 5 10 15
 Leu Leu Glu Thr Asp Ala Gly Gly Ser Val Ala Pro His Cys Ser Gly
 20 25 30
 Tyr Val Pro Trp Ser Gln Ala Leu Leu Leu Leu Arg Ser Leu Leu Glu
 35 40 45
 Met Gln Asn Leu Arg Pro Asn Ser Arg Pro Met Thr Gln Ser Leu His
 50 55 60
 Phe Asn Arg Cys Leu Cys Asp Ser Cys Ala Gly
 65 70 75

<210> 162

<211> 105

<212> PRT

<213> Homo sapiens

<400> 162

Gln Met Gln Gln Gln Asn Thr Gln Lys Val Glu Ala Ser Lys Val Pro
 1 5 10 15
 Glu Tyr Ile Lys Lys Ala Ala Lys Lys Ala Ala Glu Phe Asn Ser Asn
 20 25 30
 Leu Asn Arg Glu Arg Met Glu Glu Arg Arg Ala Tyr Phe Asp Leu Gln
 35 40 45
 Thr His Val Ile Gln Val Pro Gln Gly Lys Tyr Lys Val Leu Pro Thr
 50 55 60
 Glu Arg Thr Lys Val Ser Ser Tyr Pro Val Ala Leu Ile Pro Gly Gln
 65 70 75 80
 Phe Gln Glu Tyr Tyr Lys Ser Ile Ala Ala Phe Ala Leu His Cys Ile
 85 90 95
 Gly Tyr Trp Ala Gly Val Ser Glu Pro
 100 105

<210> 163
<211> 44
<212> PRT
<213> Homo sapiens

<400> 163
Met Thr Pro His Cys Pro Gln Asn Arg Leu His Phe Leu Leu Ala Tyr
1 5 10 15
Lys Ala Asn Leu Asn Leu Thr Pro Gly Arg His Pro Ala Thr Val Thr
20 25 30
His Ile Leu Val Ile Pro Ser Thr Ile Gly Arg Leu
35 40

<210> 164
<211> 25
<212> PRT
<213> Homo sapiens

<400> 164
Met Thr Met Trp Asn Cys Leu Leu Thr Cys Lys Val Thr His Asn Ile
1 5 10 15
Met Val Lys Phe Leu Lys Ser Asn Tyr
20 25

<210> 165
<211> 67
<212> PRT
<213> Homo sapiens

<400> 165
Met Thr Gly Tyr Cys Met Trp Glu Ile Met Lys Pro Phe Ala Val Ser
1 5 10 15
Ser Pro Val Ser Phe Arg Val Ser Val Leu Ser Lys Pro Pro Cys Glu
20 25 30
Val Asn Gln Met Leu Asp Phe Phe Pro Gln Ser His Gln Leu Pro Arg
35 40 45
Glu Arg Asp Thr Tyr Arg Thr Leu Pro Ser Ala Tyr Ser Ser Ser Ala
50 55 60
Pro Ser Thr
65

<210> 166
<211> 42
<212> PRT
<213> Homo sapiens

<400> 166

64

Met Leu Glu Met Ser Phe Ala Leu Pro Glu Phe Ala Lys Gly Ala His
 1 5 10 15

Arg Lys Gln Ile Glu Lys His Pro Leu Gly Thr Ser Leu Gln Cys Leu
 20 25 30

Leu Leu Thr Lys Phe Asn Ile Ile Asn Thr
 35 40

<210> 167

<211> 47

<212> PRT

<213> Homo sapiens

<400> 167

Met Ala Ser Val Ala Arg Lys Tyr Ala Lys Glu Glu Val Asn Pro Ile
 1 5 10 15

Ala Gly Leu Glu Asp Ser Asp Gln Thr Thr Arg Gly Leu Leu Asn Lys
 20 25 30

Gly Arg Arg Cys Pro Cys Leu Met Gly Leu Ala Trp Gly Gly Gly
 35 40 45

<210> 168

<211> 74

<212> PRT

<213> Homo sapiens

<400> 168

Met Arg Phe Ser His Phe Phe Pro Val Phe Phe Ile Thr Phe Arg Lys
 1 5 10 15

Ala Ile Leu Phe Ser Leu Tyr Thr Thr Cys Thr Leu Leu Val Gly Leu
 20 25 30

Ile Pro Arg Cys Ile Asn Ile Ile Ala Phe Met Asn Gly Ile Phe Phe
 35 40 45

Ile Val Phe Ser Asn Cys Leu Leu Asp Tyr Met Glu Ile Asp Phe Trp
 50 55 60

His Ala Asp Ile Ser Ser Lys Lys Leu Tyr
 65 70

<210> 169

<211> 27

<212> PRT

<213> Homo sapiens

<400> 169

Met Thr Lys Tyr Ser Pro Leu Pro Leu Phe Leu His Phe Ile Leu Thr
 1 5 10 15

Thr Ile Phe Phe Leu Ala Pro Phe Pro Leu Phe
 20 25

<210> 170
<211> 54
<212> PRT
<213> Homo sapiens

<400> 170
Met Leu Lys Val Arg Arg Leu Lys Asn Xaa Arg Ala Thr Val Trp Leu
1 5 10 15
Pro Gly Ile Gly Lys Gln Val Met Asp Phe Ser Leu Lys Gly Glu Ile
20 25 30
Ser Gly Val Gln Leu Gln His Leu Leu Leu Ile Asn Leu Ser Val Cys
35 40 45
Ala Ser Ser Ser Ile Glu
50

<210> 171
<211> 14
<212> PRT
<213> Homo sapiens

<400> 171
Met Pro Thr Gln Arg Gln Pro Leu Ser Ser Gln Ala Val Lys
1 5 10

<210> 172
<211> 42
<212> PRT
<213> Homo sapiens

<400> 172
Met Ala Ala Ser Val Leu Gln Ser Arg Trp Leu Ile Val Ile Leu Val
1 5 10 15
Gln Lys Arg Ile His Thr His Thr Tyr Lys Tyr Val Ser Cys Leu Asp
20 25 30
Pro Gln Glu Phe His Val Ser Leu Tyr Leu
35 40

<210> 173
<211> 121
<212> PRT
<213> Homo sapiens

<400> 173
Met Arg Thr Ser Lys Trp Ile Pro Pro Cys Lys Cys Gly Ala Gly Ala
1 5 10 15
Thr Arg His Cys Ser Gly His Ala Ser Lys Thr Gln Ala Glu Gly Ala
20 25 30

66

Ala His His Ala Gly Asp Gly Leu Lys Ala Pro Val His Ala Trp Asp
 35 40 45

Ser Ala Gln Gly Pro Cys Ser Cys Leu Gly Gln Ala Pro Gly Pro Pro
 50 55 60

Leu Ala Ala Val Ser Ser Gly Gln Gly Gly Gly Arg Tyr Gly His
 65 70 75 80

Ser Val Gly Arg Ser Trp Glu Asn Lys Ala Tyr Tyr Trp Thr Pro Gly
 85 90 95

Gly His Gly Asn His Thr Arg Met Pro Glu Thr Glu Asn Leu Trp Ala
 100 105 110

Ser Arg Ser Ser Ser Ser Cys Thr Gly
 115 120

<210> 174
 <211> 25
 <212> PRT
 <213> Homo sapiens

<400> 174
 Met Gly Asn Tyr Ala Asn Asn Lys Lys Arg Thr Leu Arg Ser Ile Asn
 1 5 10 15

Thr Val His Lys Tyr Gly Gly Leu Phe
 20 25

<210> 175
 <211> 33
 <212> PRT
 <213> Homo sapiens

<400> 175
 Met Pro Ser Phe Arg Ile Leu Asp Thr Cys Cys Phe Ser Pro Ser His
 1 5 10 15

Glu Thr Phe Cys Lys Asn Lys Glu Arg Gly Ile Thr Val Cys His His
 20 25 30

Ser

<210> 176
 <211> 30
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> (7)

<220>
 <221> UNSURE

67

<222> (11)

<400> 176

Met Ile Phe Pro Val Lys Xaa Leu Ile Arg Xaa Ile Pro Arg Asn Leu
 1 5 10 15

Leu Tyr Ile Met Asp Phe Asp Ile Tyr Leu Val Lys Val Lys
 20 25 30

<210> 177

<211> 42

<212> PRT

<213> Homo sapiens

<400> 177

Met Val Ala Ser Val Met Glu Ser Ala Asp Leu Glu Glu Gln Thr Gln
 1 5 10 15

Leu Val Thr Glu Leu Pro Gly Gly Arg Leu Ser Leu Gly Met Glu Gly
 20 25 30

Tyr Arg Asn Phe Arg Val Leu Gln Asn Phe
 35 40

<210> 178

<211> 80

<212> PRT

<213> Homo sapiens

<400> 178

Met Tyr Phe Pro Pro Ala Phe Phe Phe Pro Phe Glu Tyr Val Ser Leu
 1 5 10 15

Asn Leu Phe Ser Lys Ser Ala Arg Leu Ala Leu Ser Ser His Phe Leu
 20 25 30

Ser Leu Ser Ser Ser Tyr Leu Ser Val Phe Phe Leu Leu Val Leu Leu
 35 40 45

Phe Leu Tyr Phe Ser Pro Ser Leu His Ile His His His Lys Gln Thr
 50 55 60

Tyr Thr Phe Gln Lys Leu Val Pro Phe Trp Pro Pro Phe Asn Asn Arg
 65 70 75 80

<210> 179

<211> 40

<212> PRT

<213> Homo sapiens

<400> 179

Met Arg Val Trp Asp Pro Phe Leu Thr Leu Ile Leu Ile Lys Gln Gln
 1 5 10 15

68

Ile Phe Ile Ile Asn Glu Ile Tyr Asn Tyr Val Asn Leu Ile Asp Ile
 20 25 30

Gly Ile Val Ser Arg Ile Phe Ile
 35 40

<210> 180

<211> 82

<212> PRT

<213> Homo sapiens

<400> 180

Met Arg Tyr Thr Arg Gly Arg Arg Pro Lys Arg Arg Tyr Ile Gly His
 1 5 10 15

Leu Pro Val Phe Phe Gln Val His Phe Leu Pro Phe Ser Ala Leu Cys
 20 25 30

Tyr Asn Ser Glu Thr Asn Ile Phe Gln Leu Ser Cys Phe Leu Asp Phe
 35 40 45

Lys Lys Ala Ser Glu Arg His Cys Gly Lys Pro Lys Gly Pro Met Trp
 50 55 60

Lys Gln Ala Thr Phe His Leu Leu Arg Leu Ser Ala Ser Ser Ser Ile
 65 70 75 80

Cys Ser

<210> 181

<211> 23

<212> PRT

<213> Homo sapiens

<400> 181

Met Asp Val Ile Asp Val Pro Lys Glu Ser Val Leu Asn Leu Ile Gln
 1 5 10 15

Ser Pro Gly Ser Ser Cys Leu
 20

<210> 182

<211> 95

<212> PRT

<213> Homo sapiens

<400> 182

Met Arg Ser Ala Glu Lys Glu Arg Glu Glu Asn Thr Asn Lys Ser Leu
 1 5 10 15

Ser Ser Leu Ser Pro Val Ser Phe Pro Gln His Val Lys Gly Pro Gly
 20 25 30

Pro Lys Phe Pro Leu Pro Cys Val Leu Glu Ala Leu Leu Leu Phe Asn

69

35 40 45
 Leu Asp Thr Leu Lys Arg Glu Ala Gln Asn Thr Val Thr Val Leu Asn
 50 55 60
 Ser Lys Pro Cys His Val Thr Ser Leu His Thr Gly Leu Ala Glu Thr
 65 70 75 80
 Ser Val Gly Lys Gly Ala Ala Glu Asn Ser Val Lys Arg Lys Gln
 85 90 95

<210> 183
 <211> 31
 <212> PRT
 <213> Homo sapiens

<400> 183
 Met Arg Asn Leu Met Trp Gly Ile Arg Glu Arg Ile Lys Ser Asp Phe
 1 5 10 15
 Arg Val Phe Gly Val Ser Ile Trp Lys Ser Glu Val Ala Ile His
 20 25 30

<210> 184
 <211> 54
 <212> PRT
 <213> Homo sapiens

<400> 184
 Met Ser Phe Pro Thr Lys Gln Phe Gly Val Thr Thr Val Ile Pro Val
 1 5 10 15
 Ser Tyr Gly Trp Gly Leu Cys Ile Gly Met Cys Thr Leu Lys Phe Ile
 20 25 30
 His Leu Phe Ser Thr Ile Leu Phe Glu His Leu Leu Ser Val Arg Ala
 35 40 45
 Leu Ser Val Val Arg Tyr
 50

<210> 185
 <211> 13
 <212> PRT
 <213> Homo sapiens

<400> 185
 Met Lys Arg Glu Leu Ser Ile Leu Ile Lys Ser Lys Gly
 1 5 10

<210> 186
 <211> 51
 <212> PRT
 <213> Homo sapiens

70

<400> 186

Lys Ile Gln Ala Lys Gln Ile Lys Lys Arg Ile Gln Arg Ile Ile His
1 5 10 15

His Asp Gln Val Gly Phe Ile Pro Gly Ile Gln Gly Trp Phe Asn Ile
20 25 30

Ala Lys Ser Ile Asp Glu Thr His Lys Ile Glu Arg Ile Lys Met Arg
35 40 45

Ser Leu Met
50

<210> 187

<211> 14

<212> PRT

<213> Homo sapiens

<400> 187

Met Lys Gly Ser Tyr Leu Ile Pro Asn Phe Leu Leu Glu Pro
1 5 10

<210> 188

<211> 56

<212> PRT

<213> Homo sapiens

<400> 188

Met Asp Val Ser Ala Cys Gly Arg Leu Tyr Phe Ser Lys Met Thr Thr
1 5 10 15

Lys Ile Ser Pro Ile Ser Cys Val Ile Leu Gln Trp Gly Leu Cys Pro
20 25 30

Leu Phe Leu Asn Val Cys Ala Leu Val Thr Ala Leu Thr Asn Arg Val
35 40 45

Trp Gly Arg Met Pro Cys Asp Phe
50 55

<210> 189

<211> 29

<212> PRT

<213> Homo sapiens

<400> 189

Met Ala Leu Lys Arg Ile Val Ser His Ser Thr Arg Glu Gly Gly Thr
1 5 10 15

His Leu Glu Arg Cys His Arg Thr Pro Ile Pro Ser Gly
20 25

<210> 190

<211> 34

<212> PRT

71

<213> Homo sapiens

<400> 190

Met Thr Lys Pro Pro Ile Leu Thr Pro Trp Ser Leu Leu Ser Arg Ser
 1 5 10 15

Pro Leu Cys Ser Phe Gln Ser His Glu Glu Gly Glu Gly Arg Pro Arg
 20 25 30

Gln Gly

<210> 191

<211> 42

<212> PRT

<213> Homo sapiens

<400> 191

Met Pro Glu Ala Leu Pro Gly Pro Gly Arg Ile Lys Ser Leu Thr Val
 1 5 10 15

Trp Gly Leu Val Trp Pro Phe Thr His Ile Thr Leu Gln Asn Thr Phe
 20 25 30

Gln Gly Asp Ile Ser Val Ser Ser Ile Leu
 35 40

<210> 192

<211> 59

<212> PRT

<213> Homo sapiens

<400> 192

Met Val Gly His Lys Cys Leu Phe Asn Phe Asp Leu Leu Ala Phe Ser
 1 5 10 15

Ile Gln Ala Val Thr Leu Pro His Lys Thr Leu Gly Ala Leu Ala Arg
 20 25 30

Gly Asp Cys Thr Ser Ser Pro Gln Met Phe Ser Lys Lys Leu Pro Gly
 35 40 45

Thr Leu Leu Leu Gly Tyr Thr Lys Ser Arg Gln
 50 55

<210> 193

<211> 87

<212> PRT

<213> Homo sapiens

<400> 193

Arg Gln Cys Leu Ala Leu Ser Pro Arg Leu Glu Cys Ser Gly Thr Ile
 1 5 10 15

Ala Ala His Cys Asn Pro Arg Leu Pro Gly Ser Ser Asp Ser Tyr Ala
 20 25 30

Ser Ala Ser Arg Ala Ala Gly Ile Thr Asp Ala His Gln Asp Thr Gln
 35 40 45
 Pro Ile Phe Val Phe Leu Val Glu Met Gly Leu His His Val Cys Gln
 50 55 60
 Ala Gly Leu Glu Leu Leu Thr Ser Ser Asp Leu Pro Thr Leu Ala Ser
 65 70 75 80
 Gln Val Leu Gly Leu Gln Ala
 85

<210> 194
 <211> 117
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> (34)..(72)

<220>
 <221> UNSURE
 <222> (102)

<220>
 <221> UNSURE
 <222> (113)

<400> 194
 Met Gly Lys Ala Leu Phe Cys Gly Leu Trp Pro Leu Lys Ser Ile Cys
 1 5 10 15
 Leu Leu Leu Leu Ser Gln Gly Ser Asp Ala Ala Leu Thr Ile Leu Leu
 20 25 30
 Pro Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 35 40 45
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 50 55 60
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Leu Val Lys Cys Thr Glu Ala Cys
 65 70 75 80
 Ile Phe Glu Thr Ser Lys Gly Arg Arg Leu Arg Arg Ser Pro Leu Gln
 85 90 95
 Gly His Leu His Leu Xaa Tyr Val Ala Phe Pro Ser Asn Asn Glu Ala
 100 105 110
 Xaa His Trp Val Leu
 115

<210> 195
 <211> 47

73

<212> PRT

<213> Homo sapiens

<400> 195

Met Trp Val Ala Val Pro Asp Phe Pro Leu Leu Pro Ala Val Gly Asp
 1 5 10 15

Glu Leu Leu Ala Leu Gly Pro Asp Phe Pro Gly Trp Pro Leu Arg Ser
 20 25 30

Arg Gly Phe Lys Phe Ser Trp Ser Cys Ser Val Leu Val Gln His
 35 40 45

<210> 196

<211> 34

<212> PRT

<213> Homo sapiens

<400> 196

Met Phe Ser Leu Thr Pro Leu Glu Lys Ser Pro Ser Trp Leu Leu Ser
 1 5 10 15

Gln His Cys Pro Leu Val Ala Cys Ser Pro Trp Cys Phe Leu Ala Val
 20 25 30

Ala Thr

<210> 197

<211> 51

<212> PRT

<213> Homo sapiens

<400> 197

Met Pro Phe Pro Trp Gly Gly Leu Pro Ser Leu Ser Asn Ser Ser Leu
 1 5 10 15

Cys Trp Ser Ser Leu Pro Cys His Ser Thr Leu Ser Phe His Ser Val
 20 25 30

Cys Trp Tyr Cys Lys Tyr Leu Ile Leu Cys Ile Cys Ser Leu Ser Ala
 35 40 45

Ser Ser Gln
 50

<210> 198

<211> 286

<212> PRT

<213> Homo sapiens

<400> 198

Asn Phe Leu Glu Thr Asp Asn Glu Gly Asn Gly Ile Leu Arg Arg Arg
 1 5 10 15

Asp Ile Lys Asn Ala Leu Tyr Gly Phe Asp Ile Pro Leu Thr Pro Arg

74

20	25	30
Glu Phe Glu Lys Leu Trp Ala Arg Tyr Asp Thr Glu Gly Lys Gly His 35 40 45		
Ile Thr Tyr Gln Glu Phe Leu Gln Lys Leu Gly Ile Asn Tyr Ser Pro 50 55 60		
Ala Val His Arg Pro Cys Ala Glu Asp Tyr Phe Asn Phe Met Gly His 65 70 75 80		
Phe Thr Lys Pro Gln Gln Leu Gln Glu Glu Met Lys Glu Leu Gln Gln 85 90 95		
Ser Thr Glu Lys Ala Val Ala Ala Arg Asp Lys Leu Met Asp Arg His 100 105 110		
Gln Asp Ile Ser Lys Ala Phe Thr Lys Thr Asp Gln Ser Lys Thr Asn 115 120 125		
Tyr Ile Ser Ile Cys Lys Met Gln Glu Val Leu Glu Glu Cys Gly Cys 130 135 140		
Ser Leu Thr Glu Gly Glu Leu Thr His Leu Leu Asn Ser Trp Gly Val 145 150 155 160		
Ser Arg His Asp Asn Ala Ile Asn Tyr Leu Asp Phe Leu Arg Ala Val 165 170 175		
Glu Asn Ser Lys Ser Thr Gly Ala Gln Pro Lys Glu Lys Glu Glu Ser 180 185 190		
Met Pro Ile Asn Phe Ala Thr Leu Asn Pro Gln Glu Ala Val Arg Lys 195 200 205		
Ile Gln Glu Val Val Glu Ser Ser Gln Leu Ala Leu Ser Thr Ala Phe 210 215 220		
Ser Ala Leu Asp Lys Glu Asp Thr Gly Phe Val Lys Ala Thr Glu Phe 225 230 235 240		
Gly Gln Val Leu Lys Asp Phe Cys Tyr Lys Leu Thr Asp Asn Gln Tyr 245 250 255		
His Tyr Phe Leu Arg Lys Leu Arg Ile His Leu Thr Pro Tyr Ile Asn 260 265 270		
Trp Lys Tyr Phe Leu Gln Asn Phe Ser Cys Phe Leu Glu Glu 275 280 285		

<210> 199

<211> 64

<212> PRT

<213> Homo sapiens

<400> 199

Met Ser Gln Gln Gly Phe Phe Arg Leu Phe Gly Ile Tyr Ser Leu Pro 1 5 10 15
--

75

Ala Arg Pro Val Asn Ser Ser Arg Phe Ser Val Ser Phe Gln Ile Gly
 20 25 30
 Thr Thr Arg Asn His Gln Leu Leu Ser Tyr Thr Leu Asp Met Leu His
 35 40 45
 His Phe Asp Val Val Gly Phe Asp Tyr Tyr Lys Ile Asp Pro Asn Tyr
 50 55 60

<210> 200
 <211> 35
 <212> PRT
 <213> Homo sapiens

<400> 200
 Met Asn Lys Ile Ser Cys Phe Asn Glu Ala Asn Met Thr Ile Gln Gln
 1 5 10 15
 Cys Gly Phe Gly Ile Arg Lys Ile Leu Lys Ile Leu Ile Val Ser Phe
 20 25 30
 Ser Leu Pro
 35

<210> 201
 <211> 66
 <212> PRT
 <213> Homo sapiens

<400> 201
 Met Ser Leu Ile Leu Thr Phe His Leu Leu Leu Thr Arg Gln Ala Leu
 1 5 10 15
 Ser Pro Leu Thr Trp Ile Thr Glu Leu Thr Ser Glu Leu Gln Val Val
 20 25 30
 Ala Ser Ser Gly Pro Val Pro Ser Val Leu Phe Leu Pro Ala Arg Ile
 35 40 45
 Thr Cys Arg Ala Asp Arg Leu Phe Ala His Gly Leu His Lys Ala Ser
 50 55 60
 Arg Ala
 65

<210> 202
 <211> 27
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE

76

<222> (16)

<220>

<221> UNSURE

<222> (20)

<400> 202

Met Tyr Ala Thr Lys Lys His Val Ser Met Cys Val Asn Leu Lys Xaa
1 5 10 15

Ile Asn Gly Xaa Phe Trp Glu Val Phe Arg Ser
20 25

<210> 203

<211> 47

<212> PRT

<213> Homo sapiens

<400> 203

Met Pro Cys Leu Phe Ser Thr Ser Thr Phe Asn Phe Leu Thr Lys Ile
1 5 10 15

Lys Cys Tyr Val Phe Ser Lys Ala Asp Leu Leu Pro Ser Ser Leu Ser
20 25 30

Phe Gly Ser Ser His Tyr Gln His Ser His Pro Pro Thr Leu Lys
35 40 45

<210> 204

<211> 19

<212> PRT

<213> Homo sapiens

<400> 204

Met His Gln Ser Val Ser Leu Arg Thr Ala Trp Ala Arg His Gly Trp
1 5 10 15

Ser Arg Leu

<210> 205

<211> 22

<212> PRT

<213> Homo sapiens

<400> 205

Met Lys Ile Gln Gly Lys Asn Ile Tyr Asn Thr Thr Met Leu Lys Asp
1 5 10 15

Pro Phe Phe Tyr Leu Thr
20

<210> 206

<211> 29

<212> PRT

<213> Homo sapiens

<400> 206

Met Lys Phe His Ser Asp Pro Ser Cys Val Pro Ser Ile Gln Ile Asn
1 5 10 15

Lys Arg Asp Tyr Arg Arg Gly Pro Leu Arg Leu Ala Asn
20 25

<210> 207

<211> 21

<212> PRT

<213> Homo sapiens

<400> 207

Met Leu Pro Pro Tyr Leu Pro Lys Leu Leu Leu Gln Phe Val Phe Leu
1 5 10 15

Pro Val Ile Tyr Lys
20

<210> 208

<211> 29

<212> PRT

<213> Homo sapiens

<400> 208

Met Arg Asn Val Gln Arg Lys Phe Tyr Asn Lys Arg Val Gln Gln Gly
1 5 10 15

Cys Lys Ile Lys Asp Lys His Ile Asn Ser Ser Cys Ile
20 25

<210> 209

<211> 42

<212> PRT

<213> Homo sapiens

<400> 209

Met Glu Leu Pro Leu Phe Ser Leu Ser Cys Ser Tyr Lys Pro Cys Ala
1 5 10 15

Phe Phe Asp His Ser Thr Ala Thr Ala Ala Leu Val Met Pro Phe Leu
20 25 30

Ile Ile Pro Gly Ser His Thr Thr Arg Pro
35 40

<210> 210

<211> 18

<212> PRT

<213> Homo sapiens

<400> 210

Met Gly Tyr Leu Gly Leu Gly Met Ala Ala Gly Phe Lys Glu Arg Val

78

1 5 10 15

Val Glu

<210> 211
 <211> 70
 <212> PRT
 <213> Homo sapiens

<400> 211
 Met Glu Leu Leu Gly Ser Asp Arg Ser Pro Val Ser Phe Leu Ile His
 1 5 10 15

Trp Leu Pro Thr Arg Leu Pro His Gly Val Ser Leu Gly Ser Arg Leu
 20 25 30

Ser Ile Leu Ser Thr Phe Thr Tyr Val Asp Trp Leu Ala Glu Val Ser
 35 40 45

Thr Leu Gly Leu Asp Trp Lys Ile Leu Gln Thr Lys Lys Ala Arg Asp
 50 55 60

Ser Val Pro Pro Thr Ser
 65 70

<210> 212
 <211> 44
 <212> PRT
 <213> Homo sapiens

<400> 212
 Met Ala Asp Phe Asn Trp Met Leu Tyr Leu Gly Phe Ser Lys Ala Lys
 1 5 10 15

Lys Val Tyr Thr Leu Leu Gln Leu Gly Val Gly Leu Gln Ala Val Cys
 20 25 30

Tyr Ile His Val Leu Val Pro Val Ile Leu Thr Phe
 35 40

<210> 213
 <211> 71
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> (3)

<220>
 <221> UNSURE
 <222> (14)

<400> 213
 Met Cys Xaa Leu Gln Thr Val Tyr Ser Trp Thr Leu Leu Xaa Tyr Phe

79

1	5	10	15
Asn Pro Ser Asp	Asn Leu Cys Ile Leu Ile Arg Phe Leu Asn Pro Phe		
20	25	30	
Thr Phe Asn Val Met Phe Asp Ile Ser Trp Ile Tyr Ser Cys His Phe			
35	40	45	
Thr Phe Gly Leu Leu Cys Leu Met Tyr Phe Ser Val Leu Leu Phe Leu			
50	55	60	
Pro Tyr Cys Phe Leu Leu His			
65	70		

<210> 214
 <211> 22
 <212> PRT
 <213> Homo sapiens

<400> 214
 Met Thr Arg Ile Cys Cys Lys Ile His Phe Leu Lys Cys Leu Lys Lys
 1 5 10 15

Glu Met Glu Ile Ser Ser
 20

<210> 215
 <211> 55
 <212> PRT
 <213> Homo sapiens

<400> 215
 Met Phe Ser Met Leu Arg Tyr Cys Tyr Gln Cys Pro Leu Pro Leu Lys
 1 5 10 15

Met Thr Ala Glu Ser Lys His Phe Pro Glu Asn Ser Tyr Thr Gln Ile
 20 25 30

Phe Val Pro Leu Phe Phe Tyr Thr Ala Pro Cys Leu Phe Ile Ser Val
 35 40 45

His Ser Ser Tyr His Met Leu
 50 55

<210> 216
 <211> 49
 <212> PRT
 <213> Homo sapiens

<400> 216
 Met Pro Ser Ala Phe Glu Asn Asp Cys Arg Ile Gln Thr Phe Ser Arg
 1 5 10 15

Lys Leu Leu Tyr Ile Asp Leu Cys Ser Phe Ile Leu Leu His Ser Thr
 20 25 30

80

Leu Phe Val His Lys Cys Ser Gln Leu Ile Ser His Val Val Ile Met
 35 40 45

Cys

<210> 217
 <211> 62
 <212> PRT
 <213> Homo sapiens

<400> 217
 Met Glu Arg Cys Ala Gly Ser Glu Pro Ala Arg Lys Glu Asn Ile Ser
 1 5 10 15

Arg Leu Phe Cys Arg Met Gln Asn Trp Val Tyr Leu Gln Thr Asp Val
 20 25 30

Leu Pro Ser Lys Gly Leu Ala Thr Thr Phe Asp Pro Gln Ser Lys Val
 35 40 45

Asn Thr Ala Ile His Cys Ser Gln Thr Arg Val His Leu Pro
 50 55 60

<210> 218
 <211> 29
 <212> PRT
 <213> Homo sapiens

<400> 218
 Met Thr Thr Ser Ser Arg Thr Ile Ile Gly Lys Ile Gln Asp Leu Ser
 1 5 10 15

Val Leu Ser Thr Val Ser Gln Ile Ser Asp Arg Pro Arg
 20 25

<210> 219
 <211> 28
 <212> PRT
 <213> Homo sapiens

<400> 219
 Met Gly Phe Tyr His Lys Gly Met Ser Glu Thr Phe Ile Cys Ala Gly
 1 5 10 15

Thr Ser Ala Gln Ser Leu Asn Ala Val Ser Glu Cys
 20 25

<210> 220
 <211> 56
 <212> PRT
 <213> Homo sapiens

<400> 220
 Met Phe Ala Ser Glu Phe Phe Phe Leu Val Ile Cys Leu Val Trp Asp

81

1		5		10		15									
His	Val	Ala	Phe	Phe	Ser	Leu	Thr	Arg	Val	Ile	Lys	Val	His	Thr	Val
		20						25					30		
Lys	Ser	Met	Arg	Ser	Lys	Ala	Leu	Arg	Arg	Arg	Leu	Leu	Ser	Val	Asn
		35					40					45			
Val	Met	Ala	Gly	Ala	Ile	Arg	Leu								
	50					55									

<210> 221
 <211> 97
 <212> PRT
 <213> Homo sapiens

<400> 221
 Arg Ala Arg Ala Glu Ala Ala Arg Ala Arg Gly Glu Val Cys Phe His
 1 5 10 15
 Cys Arg Lys Pro Gly His Gly Ile Ala Asp Cys Pro Ala Ala Leu Glu
 20 25 30
 Asn Gln Asp Met Gly Thr Gly Ile Cys Tyr Arg Cys Gly Ser Thr Glu
 35 40 45
 His Glu Ile Thr Lys Cys Lys Ala Lys Val Asp Pro Ala Leu Gly Glu
 50 55 60
 Phe Pro Phe Ala Lys Cys Phe Val Cys Gly Glu Met Gly His Leu Ser
 65 70 75 80
 Arg Ser Cys Pro Asp Asn Pro Lys Gly Leu Tyr Ala Asp Gly Lys Tyr
 85 90 95

Cys

<210> 222
 <211> 36
 <212> PRT
 <213> Homo sapiens

<220>
 <221> UNSURE
 <222> (30)

<220>
 <221> UNSURE
 <222> (33)

<400> 222
 Met Ser Glu Ala Ser Leu Ser Leu Lys Glu Gln Lys Phe Cys His Pro
 1 5 10 15
 Val Val Leu Tyr Asn Leu Glu Asn Pro Leu Asn Leu Thr Xaa Leu Gln
 20 25 30

Xaa Tyr Leu Leu
35

<210> 223
<211> 65
<212> PRT
<213> Homo sapiens

<400> 223
Met Leu Cys Gly Val Leu Cys Trp Gly Trp Gly Cys Gln Asp Glu Lys
1 5 10 15
Gln Pro Cys Gly Cys Ala Leu Gly Phe Thr Ser Gln Thr Ser Val Ala
20 25 30
Phe Ala Arg Arg Lys Asp Ser Gln Gly Leu His Ile Cys Cys Pro Gln
35 40 45
Phe Cys Pro Phe Ser Asn Lys Ser His Thr Ser Asn Leu Leu Val Ala
50 55 60
His
65

<210> 224
<211> 804
<212> PRT
<213> Homo sapiens

<400> 224
Ala Lys Pro Leu Thr Asp Gln Glu Lys Arg Arg Gln Ile Ser Ile Arg
1 5 10 15
Gly Ile Val Gly Val Glu Asn Val Ala Glu Leu Lys Lys Ser Phe Asn
20 25 30
Arg His Leu His Phe Thr Leu Val Lys Asp Arg Asn Val Ala Thr Thr
35 40 45
Arg Asp Tyr Tyr Phe Ala Leu Ala His Thr Val Arg Asp His Leu Val
50 55 60
Gly Arg Trp Ile Arg Thr Gln Gln His Tyr Tyr Asp Lys Cys Pro Lys
65 70 75 80
Arg Val Tyr Tyr Leu Ser Leu Glu Phe Tyr Met Gly Arg Thr Leu Gln
85 90 95
Asn Thr Met Ile Asn Leu Gly Leu Gln Asn Ala Cys Asp Glu Ala Ile
100 105 110
Tyr Gln Leu Gly Leu Asp Ile Glu Glu Leu Glu Glu Ile Glu Glu Asp
115 120 125
Ala Gly Leu Gly Asn Gly Gly Leu Gly Arg Leu Ala Ala Cys Phe Leu
130 135 140

Asp Ser Met Ala Thr Leu Gly Leu Ala Ala Tyr Gly Tyr Gly Ile Arg
 145 150 155 160
 Tyr Glu Tyr Gly Ile Phe Asn Gln Lys Ile Arg Asp Gly Trp Gln Val
 165 170 175
 Glu Glu Ala Asp Asp Trp Leu Arg Tyr Gly Asn Pro Trp Glu Lys Ser
 180 185 190
 Arg Pro Glu Phe Met Leu Pro Val His Phe Tyr Gly Lys Val Glu His
 195 200 205
 Thr Asn Thr Gly Thr Lys Trp Ile Asp Thr Gln Val Val Leu Ala Leu
 210 215 220
 Pro Tyr Asp Thr Pro Val Pro Gly Tyr Met Asn Asn Thr Val Asn Thr
 225 230 235 240
 Met Arg Leu Trp Ser Ala Arg Ala Pro Asn Asp Phe Asn Leu Arg Asp
 245 250 255
 Phe Asn Val Gly Asp Tyr Ile Gln Ala Val Leu Asp Arg Asn Leu Ala
 260 265 270
 Glu Asn Ile Ser Arg Val Leu Tyr Pro Asn Asp Asn Val Ala Ile Gln
 275 280 285
 Leu Asn Asp Thr His Pro Ala Leu Ala Ile Pro Glu Leu Met Arg Ile
 290 295 300
 Phe Val Asp Ile Glu Lys Leu Pro Trp Ser Lys Ala Trp Glu Leu Thr
 305 310 315 320
 Gln Lys Thr Phe Ala Tyr Thr Asn His Thr Val Leu Pro Glu Ala Leu
 325 330 335
 Glu Arg Trp Pro Val Asp Leu Val Glu Lys Leu Leu Pro Arg His Leu
 340 345 350
 Glu Ile Ile Tyr Glu Ile Asn Gln Lys His Leu Asp Arg Ile Val Ala
 355 360 365
 Leu Phe Pro Lys Asp Val Asp Arg Leu Arg Arg Met Ser Leu Ile Glu
 370 375 380
 Glu Glu Gly Ser Lys Arg Ile Asn Met Ala His Leu Cys Ile Val Gly
 385 390 395 400
 Ser His Ala Val Asn Gly Val Ala Lys Ile His Ser Asp Ile Val Lys
 405 410 415
 Thr Lys Val Phe Lys Asp Phe Ser Glu Leu Glu Pro Asp Lys Phe Gln
 420 425 430
 Asn Lys Thr Asn Gly Ile Thr Pro Arg Arg Trp Leu Leu Leu Cys Asn
 435 440 445
 Pro Gly Leu Ala Glu Leu Ile Ala Glu Lys Ile Gly Glu Asp Tyr Val

450	455	460
Lys Asp Leu Ser Gln Leu Thr Lys Leu His Ser Phe Leu Gly Asp Asp		
465	470	475 480
Val Phe Leu Arg Glu Leu Ala Lys Val Lys Gln Glu Asn Lys Leu Lys		
	485	490 495
Phe Ser Gln Phe Leu Glu Thr Glu Tyr Lys Val Lys Ile Asn Pro Ser		
	500	505 510
Ser Met Phe Asp Val Gln Val Lys Arg Ile His Glu Tyr Lys Arg Gln		
	515	520 525
Leu Leu Asn Cys Leu His Val Ile Thr Met Tyr Asn Arg Ile Lys Lys		
	530	535 540
Asp Pro Lys Lys Leu Phe Val Pro Arg Thr Val Ile Ile Gly Gly Lys		
	545	550 555 560
Ala Ala Pro Gly Tyr His Met Ala Lys Met Ile Ile Lys Leu Ile Thr		
	565	570 575
Ser Val Ala Asp Val Val Asn Asn Asp Pro Met Val Gly Ser Lys Leu		
	580	585 590
Lys Val Ile Phe Leu Glu Asn Tyr Arg Val Ser Leu Ala Glu Lys Val		
	595	600 605
Ile Pro Ala Thr Asp Leu Ser Glu Gln Ile Ser Thr Ala Gly Thr Glu		
	610	615 620
Ala Ser Gly Thr Gly Asn Met Lys Phe Met Leu Asn Gly Ala Leu Thr		
	625	630 635 640
Ile Gly Thr Met Asp Gly Ala Asn Val Glu Met Ala Glu Glu Ala Gly		
	645	650 655
Glu Glu Asn Leu Phe Ile Phe Gly Met Arg Ile Asp Asp Val Ala Ala		
	660	665 670
Leu Asp Lys Lys Gly Tyr Glu Ala Lys Glu Tyr Tyr Glu Ala Leu Pro		
	675	680 685
Glu Leu Lys Leu Val Ile Asp Gln Ile Asp Asn Gly Phe Phe Ser Pro		
	690	695 700
Lys Gln Pro Asp Leu Phe Lys Asp Ile Ile Asn Met Leu Phe Tyr His		
	705	710 715 720
Asp Arg Phe Lys Val Phe Ala Asp Tyr Glu Ala Tyr Val Lys Cys Gln		
	725	730 735
Asp Lys Val Ser Gln Leu Tyr Met Asn Pro Lys Ala Trp Asn Thr Met		
	740	745 750
Val Leu Lys Asn Ile Ala Ala Ser Gly Lys Phe Ser Ser Asp Arg Thr		
	755	760 765

85

Ile Lys Glu Tyr Ala Gln Asn Ile Trp Asn Val Glu Pro Ser Asp Leu
 770 775 780

Lys Ile Ser Leu Ser Asn Glu Ser Asn Lys Val Asn Gly Asn Asn Lys
 785 790 795 800

Val Asn Gly Asn

<210> 225

<211> 60

<212> PRT

<213> Homo sapiens

<400> 225

Met Gly Asp Leu Tyr Lys Lys Glu Leu Lys Lys Arg Arg Asn Val Ile
 1 5 10 15

Ser Met Leu Leu Gln Val Lys Gly Lys Gln Glu Asp Lys Tyr His Lys
 20 25 30

Lys Thr Lys Met Tyr Leu Thr Phe Trp Asp Lys Ile Val Gly Ser Thr
 35 40 45

Glu Asn Trp Asn Leu Glu Leu Pro Val Pro Gln Arg
 50 55 60

<210> 226

<211> 46

<212> PRT

<213> Homo sapiens

<400> 226

Met Phe Tyr Glu Tyr Lys Glu Tyr Asn Glu Cys Tyr Tyr Lys Tyr Ile
 1 5 10 15

His Ala Asn Arg Asp Phe Gln Tyr Pro Thr Phe Ser Gln Phe Arg Leu
 20 25 30

Pro Glu Ile Gly Leu Leu Gly Gln Arg Leu Gln Thr Tyr Phe
 35 40 45

<210> 227

<211> 13

<212> PRT

<213> Homo sapiens

<400> 227

Met Arg Arg Trp Tyr Ile Trp Glu Val Ser Arg Gly Tyr
 1 5 10

<210> 228

<211> 27

<212> PRT

<213> Homo sapiens

<400> 228

Met Phe Leu Arg Tyr Leu Gly Lys Ser Ser Glu Pro Cys Val Ala Asn
1 5 10 15

Gly Asn Ala Val Val Gln Trp Gly Leu Leu Gly
20 25

<210> 229

<211> 45

<212> PRT

<213> Homo sapiens

<400> 229

Met Ala Thr Asn Ser Cys Leu Tyr Ser Thr His Lys Gln Phe Gln Tyr
1 5 10 15

Met Phe Cys Asp Arg Ser Pro Lys Ile Ser Ser Phe Met Val Pro Gly
20 25 30

Arg Thr Glu Asn Ser Arg Met Gln Leu Leu Lys Leu Phe
35 40 45

<210> 230

<211> 96

<212> PRT

<213> Homo sapiens

<400> 230

Lys Arg Gln Gly Leu Ala Leu Ser Pro Arg Leu Glu Tyr Asn Asp Val
1 5 10 15

Ile Ile Ala His Arg Asn Phe Glu Leu Pro Gly Ser Ser Asn Pro Ser
20 25 30

Ala Ser Ala Ser Gln Glu Leu Gly Leu Gln Thr Cys Ala Thr Thr Ser
35 40 45

Ser Phe Phe Ile Phe Cys Arg Gly Arg Val Ser Leu Cys Cys Pro Gly
50 55 60

Gly Val Ser His Ser Thr Ser Ser Asn Pro Thr Ala Ser Ala Ser Gln
65 70 75 80

Arg Ala Arg Ile Thr Gly Leu Ser His Cys Thr Gln Pro Lys Ala Leu
85 90 95

<210> 231

<211> 56

<212> PRT

<213> Homo sapiens

<400> 231

87

Met Leu Ala Leu Ser His Trp Thr Val Val Pro Ser His Pro Leu Ser
 1 5 10 15

Pro Ser Leu Asp His Glu His Ser Arg Ala Arg Thr Thr Ser Val Leu
 20 25 30

Phe Thr Ala Val His Pro Ala Leu Thr Gln Cys Leu Met His Ala Leu
 35 40 45

Gly Ala Gln Glu Val Leu Ile Gln
 50 55

<210> 232
 <211> 34
 <212> PRT
 <213> Homo sapiens

<400> 232
 Met Asp Ser Pro Lys Arg Val Ser Ser Asp Leu Ser Leu Leu Arg Asn
 1 5 10 15

Lys Ile Leu Asp Ser Gly Cys Val Cys Phe Arg Cys Cys Gly Thr Gly
 20 25 30

Trp Phe

<210> 233
 <211> 34
 <212> PRT
 <213> Homo sapiens

<400> 233
 Met Leu Ser Ala Phe Phe Thr Leu Ile Leu Ser Pro Val Tyr Arg Arg
 1 5 10 15

Val Phe Gln Arg Leu His Met Arg Tyr Leu Asn Lys Leu Lys Ala Glu
 20 25 30

Glu Ile

<210> 234
 <211> 35
 <212> PRT
 <213> Homo sapiens

<400> 234
 Met Cys Phe Glu Thr Gly Glu Tyr Ser Trp Ser Gly Ala Gly Ala Gln
 1 5 10 15

Asn Thr Arg Phe Leu Cys Ser Asp Asn Leu Cys Ser Leu Ala Leu Leu
 20 25 30

Leu Ile Tyr
 35

<210> 235
<211> 40
<212> PRT
<213> Homo sapiens

<400> 235
Met Ile Asn Glu Gln Met Asn Ile Ser Glu Lys Leu Val Tyr Ile Ile
1 5 10 15
Met Asn Arg Leu Val Leu His Phe Tyr Lys Asn Arg Lys Leu Lys Ile
20 25 30
Lys Lys Lys Ile Leu Pro Lys Lys
35 40

<210> 236
<211> 60
<212> PRT
<213> Homo sapiens

<400> 236
Met Tyr Lys Cys Leu Leu Glu Ala His Glu Val Tyr Arg Trp Phe Leu
1 5 10 15
Pro Gln Tyr Leu Thr Ile Val Lys Phe Gln Ala Met Pro Leu Leu Ser
20 25 30
Thr Thr Phe Ser Leu Arg Ser Thr Gly Ile Trp Leu Arg Phe His Ser
35 40 45
Asp Asp Leu Leu Ser Glu Thr Leu Arg Leu Glu Lys
50 55 60

<210> 237
<211> 36
<212> PRT
<213> Homo sapiens

<400> 237
Met Ser Leu Tyr Leu Phe Ser Pro Phe His Cys Pro Phe Phe Phe Pro
1 5 10 15
His Leu Pro Leu Cys Ser Val Leu Ser Leu Ala Ser Ser Cys Gln Tyr
20 25 30
Val Asp Phe Cys
35

<210> 238
<211> 66
<212> PRT
<213> Homo sapiens

<400> 238

89

Met Phe Phe Tyr Leu Ser Lys Thr Leu Pro Met Phe Leu Leu Lys His
 1 5 10 15

His Ser Tyr Ser Lys Thr Lys Val Asn Glu Asn Leu Tyr Gln Asp Asp
 20 25 30

Cys Pro Gln Ser Ser Gly Trp Thr Thr Cys Leu Ser Ser Ile Ile Leu
 35 40 45

Cys Ile Ile Ser Leu Ile His Ser Asn Ser Leu Cys Ile Ile Cys Ala
 50 55 60

Ser Gly
 65

<210> 239

<211> 31

<212> PRT

<213> Homo sapiens

<400> 239

Met Cys His Gly Phe Val Thr Pro Tyr Tyr Tyr Tyr Leu Ser Leu Ala
 1 5 10 15

Ser Cys Tyr Cys Pro Tyr Leu Thr Thr Ile Thr Ser Met Ser Ser
 20 25 30

<210> 240

<211> 44

<212> PRT

<213> Homo sapiens

<400> 240

Met Asn Asn Ile Ile Pro Leu Leu Ile Leu Met Gly Leu Phe Phe Leu
 1 5 10 15

Ser Gln Ser Ala Leu Ile His Ile Gly Ser Leu Asn Ser Ser Asn Ile
 20 25 30

Ile Lys Ser Phe Ser Pro Arg Asp Pro Thr Phe Arg
 35 40

THIS PAGE BLANK (USPTO)